

STIC-Biotech/ChemLib

70812

From: Sullivan, Daniel  
Sent: Sunday, July 14, 2002 10:52 AM  
To: STIC-Biotech/ChemLib  
Subject: CDB Search Request (10/053641)

For case number **10/053641**, please do a standard sequence search for SEQ ID NO: 1-8 against the commercial or interference nucleic acid databases or both.

Thanks very much.

Daniel M. Sullivan  
Examiner AU 1636  
Room: 12D12  
Mail Box: 11E12  
Tel: 703-305-4448

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 7/15/02  
Date Completed: 7/16/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: 8  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 01  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

2.12



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 17:12:34 ; Search time 2368.24 Seconds  
(without alignments)  
918.978 Million cell updates/sec

Title: US-10-053-641-1

Perfect score: 104

Sequence: 1 gatcctttatgtgtttac.....gcacaaatgcatcctgggc 104

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
-----						

1	100	96.2	226	12	SYNHIR	M26762 Synthetic h
2	96	92.3	129	6	I41183	I41183 Sequence 3
3	96	92.3	235	6	I41187	I41187 Sequence 7
c 4	96	92.3	235	6	I41188	I41188 Sequence 8
5	96	92.3	726	6	I41189	I41189 Sequence 9
c 6	96	92.3	726	6	I41190	I41190 Sequence 10
7	88.4	85.0	220	12	SYNHIRMED	M14964 Synthetic h
8	88.4	85.0	224	6	A00630	A00630 Artificial
c 9	88.4	85.0	224	6	A00631	A00631 Artificial
10	88.4	85.0	224	6	A04277	A04277 Artificial
c 11	88.4	85.0	224	6	A04278	A04278 Artificial
12	88.4	85.0	224	6	A07345	A07345 Synthetic D
c 13	88.4	85.0	224	6	A07346	A07346 Synthetic D
14	86.8	83.5	109	6	A34234	A34234 Synthetic h
c 15	86.8	83.5	109	6	A34235	A34235 Synthetic h
16	86.8	83.5	217	6	A03695	A03695 Nucleotide
c 17	86.8	83.5	217	6	A03696	A03696 Nucleotide
18	86.8	83.5	217	6	A34238	A34238 Synthetic d
c 19	86.8	83.5	217	6	A34239	A34239 Synthetic d
20	86.8	83.5	217	6	E00657	E00657 DNA encodin
c 21	83.2	80.0	178	6	A13978	A13978 Synthetic D
22	83.2	80.0	180	6	A02448	A02448 Nucleotide
c 23	83.2	80.0	180	6	A02449	A02449 Nucleotide
24	83.2	80.0	180	6	A13977	A13977 Synthetic D
c 25	82.4	79.2	195	6	A03693	A03693 Nucleotide
26	82.4	79.2	195	6	A03694	A03694 Nucleotide
c 27	82.4	79.2	279	6	A01139	A01139 Fuson DNA
28	82.4	79.2	279	6	I26634	I26634 Sequence 3
c 29	82.2	79.9	238	6	A04616	A04616 HVI gene. 8
30	81	77.9	106	6	I41186	I41186 Sequence 6
c 31	80.8	77.7	225	6	A04613	A04613 Nucleotide
32	80.8	77.7	225	6	A04614	A04614 Nucleotide
c 33	80.8	77.7	226	6	A04615	A04615 Nucleotide
34	79	76.0	93	6	I63540	I63540 Sequence 3
c 35	78.2	75.2	8491	6	AR031529	AR031529 Sequence
36	78	75.0	212	6	A13382	A13382 DNA sequenc
c 37	78	75.0	212	6	A13383	A13383 DNA sequenc
38	78	75.0	212	6	A34619	A34619 Synthetic h
c 39	78	75.0	212	6	A34620	A34620 Synthetic h
40	78	75.0	212	6	E00711	E00711 Synthetic D
41	77.4	74.4	4477	12	SYNEXVECHD	M88335 Expression
42	74.2	71.3	223	6	A19994	A19994 SEQ ID NO:
c 43	74.2	71.3	223	6	A20035	A20035 SEQ ID NO:
44	72.6	69.8	211	6	E03003	E03003 DNA encodin
45	72	69.2	1467	6	A20030	A20030 SEQ ID NO:

## ALIGNMENTS

RESULT	1	SYNHIR	SYNHIR	226 bp	DNA	linear	SYN 27-APR-1993
LOCUS	SYNHIR	Synthetic hirudin gene, complete cds.					
DEFINITION	1	(bases 1 to 226)					
ACCESSION	M26762						
VERSION	M26762.1	GI:208478					
KEYWORDS	hirudin; proteinase inhibitor; thrombin inhibitor.						
SOURCE	Synthetic DNA.						
ORGANISM	artificial construct						
REFERENCE	1	(bases 1 to 226)					
AUTHORS	Bergmann,C., Dodd,J., Koehler,S., Fink,E. and Gassen,H.G.						
TITLE	Chemical synthesis and expression of a gene coding for hirudin, the thrombin-specific inhibitor from the leech Hirudo medicinalis						
JOURNAL	Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)						
MEDLINE	87026239						
FEATURES	source	Location/Qualifiers					
		1..226					
		/organism="synthetic construct"					
		/db_xref="taxon:32630"					
		22..222					
		/note="hirudin"					
		/codon_start=1					

/transl\_table=11  
/protein\_id="AA72772.1"  
/db\_xref="GI:208479"  
/translation="MVYTDCTESGNLCLCEGNNVCGGNKILGSDGKNOCVTGE  
GTPRPQHNDGDFEIPPEYLQ"  
BASE COUNT 63 a 56 c 61 g 46 t  
ORIGIN

Query Match 96.2%; Score 100; DB 12; Length 226;  
Best Local Similarity 100.0%; Pred. No. 1e-22;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 cttatgtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaagg 64  
|||||  
Db 18 CTTTATGTTTACACTGACTGACTGAATCCGGTCAGAACCTGTGCTGTGCGAAGG 77  
|||||

QY 65 ctctaactgttcgagcagggaacaatgcatcctgggc 104  
|||||

Db 78 CTCFAACGTTTGGGCCAGGCAACAATGATCTCTGGGC 117  
|||||

RESULT 2  
LOCUS I41183 129 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 3 from patent US 5624822.  
ACCESSION I41183  
VERSION I41183.1 GI:2081773  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 129)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 3 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..129

BASE COUNT 36 a 35 c 29 g 29 t  
ORIGIN

Query Match 92.3%; Score 96; DB 6; Length 129;  
Best Local Similarity 100.0%; Pred. No. 2.1e-21;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggctct 68  
|||||  
Db 31 ATGGTTGTTTACACTGACTGACTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 90  
|||||

QY 69 aacgtttgcgagcagggaacaatgcatcctgggc 104  
|||||  
Db 91 AACGTTTGGCGCCAGGCAACAATGATGATCTCTGGGC 126  
|||||

RESULT 3  
LOCUS I41187 235 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 7 from patent US 5624822.  
ACCESSION I41187  
VERSION I41187.1 GI:2081777  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 235)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 7 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..235  
/organism="unknown"

BASE COUNT  
ORIGIN

72 a 63 c 57 g 43 t

Query Match 92.3%; Score 96; DB 6; Length 235;  
Best Local Similarity 100.0%; Pred. No. 2.2e-21;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggctct 68  
|||||  
Db 31 ATGGTTGTTTACACTGACTGACTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 90  
|||||

QY 69 aacgtttgcgagcagggaacaatgcatcctgggc 104  
|||||  
Db 91 AACGTTTGGCGCCAGGCAACAATGATGATCTCTGGGC 126  
|||||

RESULT 4  
LOCUS I41188 235 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 8 from patent US 5624822.  
ACCESSION I41188  
VERSION I41188.1 GI:2081778  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 235)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 8 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..235  
/organism="unknown"

BASE COUNT 42 a 58 c 64 g 71 t  
ORIGIN

Query Match 92.3%; Score 96; DB 6; Length 235;  
Best Local Similarity 100.0%; Pred. No. 2.2e-21;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggctct 68  
|||||  
Db 209 ATGGTTGTTTACACTGACTGACTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 150  
|||||

QY 69 aacgtttgcgagcagggaacaatgcatcctgggc 104  
|||||  
Db 149 AACGTTTGGCGCCAGGCAACAATGATGATCTCTGGGC 114  
|||||

RESULT 5  
LOCUS I41189 726 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 9 from patent US 5624822.  
ACCESSION I41189  
VERSION I41189.1 GI:2081779  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 726)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 9 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..726  
/organism="unknown"

BASE COUNT 281 a 171 c 126 g 148 t  
ORIGIN

Query Match 92.3%; Score 96; DB 6; Length 726;

```
Best Local Similarity 100.0%; Pred. No. 2.3e-21;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggtgttttaactgactgactgaatccggtcagaacctgtgctgtgcgaaggctct 68
|||||
Db 523 ATGGTTGTTTACACTGACTGCTGAATCCGCTCAGAACCTGTGCTGTGCGAAGGCTCT 582
|||||

QY 69 aacgtttgcggcaggcgaacaaatgcatcctgggc 104
|||||
Db 583 AACGTTTGGCGGCAGGCAACAATGCAATCCTCTGGGC 618
|||||

RESULT 6
LOCUS I41190/c 141190 726 bp DNA linear PAT 13-MAY-1997
DEFINITION Sequence 10 from patent US 5624822.
ACCESSION I41190
VERSION I41190.1 GI:2081780
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 726)
AUTHORS Koerwer, W.
TITLE Hirudin fusion proteins and preparation of hirudin
JOURNAL Patent: US 5624822-A 10 29-APR-1997;
FEATURES
source
BASE COUNT 149 a 126 c 171 g 280 t
ORIGIN

Query Match 92.3%; Score 96; DB 6; Length 726;
Best Local Similarity 100.0%; Pred. No. 2.3e-21;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atggtgttttaactgactgactgaatccggtcagaacctgtgctgtgcgaaggctct 68
|||||
Db 204 ATGGTTGTTTACACTGACTGCTGAATCCGCTCAGAACCTGTGCTGTGCGAAGGCTCT 145
|||||

QY 69 aacgtttgcggcaggcgaacaaatgcatcctgggc 104
|||||
Db 144 AACGTTTGGCGGCAGGCAACAATGCAATCCTCTGGGC 109
|||||

RESULT 7
LOCUS SYNHIEMED 220 bp DNA linear SYN 27-APR-1993
DEFINITION Synthetic hirudin gene, 3' end.
ACCESSION M14964
VERSION M14964.1 GI:208480
KEYWORDS hirudin.
SOURCE Synthetic DNA.
ORGANISM Synthetic construct.
REFERENCE 1 (bases 1 to 220)
AUTHORS Fortkamp, E., Rieger, M., Heisterberg-Moutsers, G., Schweitzer, S. and Sommer, R.
TITLE Cloning and expression in Escherichia coli of a synthetic DNA for hirudin, the blood coagulation inhibitor in the leech
JOURNAL DNA 5, 511-517 (1986)
MEDLINE 87132924
COMMENT Draft entry and printed copy of sequence for [1] kindly provided by M. Rieger, 23-NOV-1988.
FEATURES
source
location/Qualifiers
1..220
/organism="synthetic construct"
/db_xref="taxon:32630"
<1..207
/note="hirudin precursor"
/codon_start=1
/transl_table=11

CDS
/organism="synthetic construct"
/db_xref="taxon:32630"
<1..207
/codon_start=1
/transl_table=11
/product="hirudin-like protein"
/protein_id="CAA00072.1"
/db_xref="GI:14561"
/translation="NSMVTYDCTESQNLCLCEGSNVCGQGNKILGSDGEKNOCVT
GEGTPKQSHNDGDFEIPPEYLQ"
BASE COUNT 62 a 51 c 60 g 51 t
ORIGIN

Query Match 85.0%; Score 88.4; DB 6; Length 224;
Best Local Similarity 93.9%; Pred. No. 7.4e-19;
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 6 tttatggtgtttacactgactgactgaatcccggtcagaacctgtgctgtgcgaaggc 65
|||||
Db 4 TCTATGTTGTTTACACTGACTGCACCGAATCTGTCAGAACCTGTGCTGTGCGAAGGT 63
|||||

QY 66 tctaactgttgcggcaggcgaacaaatgcatcctggg 103
|||||
Db 64 TCTAACGTTTGGCTCAGGGAACAATGCAATCCTCTGGG 101
|||||

RESULT 8
LOCUS A00630 224 bp DNA linear PAT 06-JUL-1995
DEFINITION Artificial sequence for hirudin-like protein.
ACCESSION A00630
VERSION A00630.1 GI:14560
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 224)
AUTHORS
TITLE DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR PREPARING SUCH PROTEIN
JOURNAL Patent: WO 8603517-A 15 19-JUN-1986;
FEATURES
source
location/Qualifiers
1..224
/organism="synthetic construct"
/db_xref="taxon:32630"
<1..207
/codon_start=1
/transl_table=11
/product="hirudin-like protein"
/protein_id="CAA00072.1"
/db_xref="GI:14561"
/translation="NSMVTYDCTESQNLCLCEGSNVCGQGNKILGSDGEKNOCVT
GEGTPKQSHNDGDFEIPPEYLQ"
BASE COUNT 62 a 51 c 60 g 51 t
ORIGIN

Query Match 85.0%; Score 88.4; DB 6; Length 224;
Best Local Similarity 93.9%; Pred. No. 7.4e-19;
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 6 tttatggtgtttacactgactgactgaatcccggtcagaacctgtgctgtgcgaaggc 65
|||||
Db 4 TCTATGTTGTTTACACTGACTGCACCGAATCTGTCAGAACCTGTGCTGTGCGAAGGT 63
|||||

QY 66 tctaactgttgcggcaggcgaacaaatgcatcctggg 103
|||||
Db 64 TCTAACGTTTGGCTCAGGGAACAATGCAATCCTCTGGG 101
|||||

RESULT 9
LOCUS A00631/c 224 bp DNA linear PAT 30-MAR-1993
DEFINITION Artificial sequence for hirudin-like protein, reverse complement.
ACCESSION A00631
VERSION A00631.1 GI:14562
```

KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 224)  
AUTHORS  
TITLE DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR  
PREPARING SUCH PROTEIN  
JOURNAL Patent: WO 8603517-A 16 JUN-1986;  
FEATURES Location/Qualifiers  
source  
1..224  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 51 a 60 c 51 g 62 t  
ORIGIN

Query Match 85.0%; Score 88.4; DB 6; Length 224;  
Best Local Similarity 93.9%; Pred. No. 7.4e-19;  
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 6 ttatggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaagc 65  
Db 221 TCTATGTTGTTTACACTGACTGCACCGAATCTGTCAGAACCTGTGCCTGTGCGAAGT 162  
Qy 66 tctaacgtttgcccaggcgcaacaaatgcctcctggg 103  
Db 161 TCTAACGTTTGGGTCAGGGAAACAAATGCATCCTGGG 124

RESULT 10  
LOCUS A04277 224 bp DNA linear PAT 15-APR-1993  
DEFINITION Artificial sequence for hirudin-like fusion protein.  
ACCESSION A04277  
VERSION A04277.1 GI:344867  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 224)  
AUTHORS  
TITLE DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR  
PREPARING SUCH PROTEIN  
JOURNAL Patent: WO 8603517-A 19 JUN-1986;  
FEATURES Location/Qualifiers  
source  
1..224  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 62 a 51 c 60 g 51 t  
ORIGIN

Query Match 85.0%; Score 88.4; DB 6; Length 224;  
Best Local Similarity 93.9%; Pred. No. 7.4e-19;  
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 6 ttatggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaagc 65  
Db 4 TCTATGTTGTTTACACTGACTGCACCGAATCTGTCAGAACCTGTGCCTGTGCGAAGT 63  
Qy 66 tctaacgtttgcccaggcgcaacaaatgcctcctggg 103  
Db 64 TCTAACGTTTGGGTCAGGGAAACAAATGCATCCTGGG 101

RESULT 11  
LOCUS A04278/c 224 bp DNA linear PAT 15-APR-1993  
DEFINITION Artificial sequence for hirudin-like fusion protein, reverse  
complement.  
ACCESSION A04278  
VERSION A04278.1 GI:344868

KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 224)  
AUTHORS  
TITLE DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR  
PREPARING SUCH PROTEIN  
JOURNAL Patent: WO 8603517-A 20 JUN-1986;  
FEATURES Location/Qualifiers  
source  
1..224  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 51 a 60 c 51 g 62 t  
ORIGIN

Query Match 85.0%; Score 88.4; DB 6; Length 224;  
Best Local Similarity 93.9%; Pred. No. 7.4e-19;  
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 6 ttatggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaagc 65  
Db 221 TCTATGTTGTTTACACTGACTGCACCGAATCTGTCAGAACCTGTGCCTGTGCGAAGT 162  
Qy 66 tctaacgtttgcccaggcgcaacaaatgcctcctggg 103  
Db 161 TCTAACGTTTGGGTCAGGGAAACAAATGCATCCTGGG 124

RESULT 12  
LOCUS A07345 224 bp DNA linear PAT 03-AUG-1993  
DEFINITION Synthetic DNA for hirudin-similar protein.  
ACCESSION A07345  
VERSION A07345.1 GI:413049  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 224)  
AUTHORS  
JOURNAL Patent: DE 3445517-A 1 19-JUN-1986;  
FEATURES Location/Qualifiers  
source  
1..224  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 62 a 51 c 60 g 51 t  
ORIGIN

Query Match 85.0%; Score 88.4; DB 6; Length 224;  
Best Local Similarity 93.9%; Pred. No. 7.4e-19;  
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 6 ttatggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaagc 65  
Db 4 TCTATGTTGTTTACACTGACTGCACCGAATCTGTCAGAACCTGTGCCTGTGCGAAGT 63  
Qy 66 tctaacgtttgcccaggcgcaacaaatgcctcctggg 103  
Db 64 TCTAACGTTTGGGTCAGGGAAACAAATGCATCCTGGG 101

RESULT 13  
LOCUS A07346/c 224 bp DNA linear PAT 03-AUG-1993  
DEFINITION Synthetic DNA for hirudin-similar protein  
ACCESSION A07346  
VERSION A07346.1 GI:411427  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct

artificial sequence.  
REFERENCE 1 (bases 1 to 224)  
AUTHORS  
JOURNAL  
FEATURES source  
Patent: DE 3445517-A 2 19-JUN-1986;  
Location/Qualifiers  
1. .224  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 51 a 60 c 51 g 62 t  
ORIGIN

Query Match 85.0%; Score 88.4; DB 6; Length 224;  
Best Local Similarity 93.9%; Pred. No. 7.4e-19;  
Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 6 ttatggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaagc 65  
DB 221 TCTATGTTGTTTACACTGACTGCACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGGT 162  
QY 66 tctaactgtttgcggcagggaacaaatgcatcctggg 103  
DB 161 TCTAACGTTTGGCGTCAGGGAACAAATGCATCCTGGG 124

RESULT 14  
A34234  
LOCUS A34234 109 bp DNA linear PAT 25-JUL-1996  
DEFINITION Synthetic hirudin analogue duplex I.  
ACCESSION A34234  
VERSION A34234.1 GI:1568164  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 109)  
AUTHORS Liersch, M., Rink, H., Maerkl, W., Gruetter, M.G. and Meyhack, B.  
TITLE Process for the preparation of thrombin inhibitors  
JOURNAL Patent: EP 0168342-A 6 15-JAN-1986;  
CIBA-GEIGY AG; CIBA-GEIGY AG; UCP GEN-PHARMA AG  
FEATURES source  
1. .109  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 23 a 25 c 30 g 31 t  
ORIGIN

Query Match 83.5%; Score 86.8; DB 6; Length 109;  
Best Local Similarity 92.9%; Pred. No. 2.4e-18;  
Matches 91; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 6 ttatggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaagc 65  
DB 7 TTCATGTTGTTTACACCGACTGCACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGGT 66  
QY 66 tctaactgtttgcggcagggaacaaatgcatcctggg 103  
DB 67 TCTAACGTTTGGCGTCAGGGAACAAATGCATCCTGGG 104

RESULT 15  
A34235/C  
LOCUS A34235 109 bp DNA linear PAT 25-JUL-1996  
DEFINITION Synthetic hirudin analogue duplex I.  
ACCESSION A34235  
VERSION A34235.1 GI:1568165  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 109)  
AUTHORS Liersch, M., Rink, H., Maerkl, W., Gruetter, M.G. and Meyhack, B.

TITLE Process for the preparation of thrombin inhibitors  
JOURNAL Patent: EP 0168342-A 7 15-JAN-1986;  
CIBA-GEIGY AG; CIBA-GEIGY AG; UCP GEN-PHARMA AG  
FEATURES source  
1. .109  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 31 a 30 c 25 g 23 t  
ORIGIN

Query Match 83.5%; Score 86.8; DB 6; Length 109;  
Best Local Similarity 92.9%; Pred. No. 2.4e-18;  
Matches 91; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 6 ttatggtgtttacactgactgactgaatccggtcagaacctgtgcctgtgcgaagc 65  
DB 103 TTCATGTTGTTTACACCGACTGCACCGAATCTGGTCAGAACCTGTGCCTGTGCGAAGGT 44  
QY 66 tctaactgtttgcggcagggaacaaatgcatcctggg 103  
DB 43 TCTAACGTTTGGCGTCAGGGAACAAATGCATCCTGGG 6

Search completed: July 15, 2002, 22:28:32  
Job time: 18958 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 19:57:07 ; Search time 758.37 seconds  
(without alignments)  
235.451 Million cell updates/sec

Title: US-10-053-641-1

Perfect score: 104

Sequence: 1 gatcctttatgttggtttac.....gcacaaatgcatcctgggc 104

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-nb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	ID	Description
1	96	92.3	726 12	AAQ12379 Hirudin peptide/Pr
2	93	89.4	468 16	AAQ93244 Fusion construct o
3	88.4	85.0	224 7	AAN60746 DNA encoding hirud
4	86.8	83.5	217 7	AAN60355 Desulphatohirudin
5	86.8	83.5	217 8	AAN70323 Sequence of the de
6	82.4	79.2	196 8	AAN70319 Sequence encoding
7	82.4	79.2	279 12	AAQ13831 MSP signal peptide
8	80.8	77.7	229 9	AAN81294 Sequence encoding
9	79	76.0	238 9	AAN81295 Sequence encoding

10	78.2	75.2	227	22	AAF61507	S. marcescens hiru
11	72.6	69.8	210	12	AAQ10452	HV-1 gene. Synthe
12	72	69.2	1467	12	AAQ12490	Factor xa-cleavabl
13	71.2	68.5	201	12	AAQ12153	Synthetic hirudin
14	71.2	68.5	201	13	AAQ25184	Hirudin HV-1. Syn
15	71.2	68.5	420	12	AAQ12155	Factor xa-cleavabl
16	71.2	68.5	1458	12	AAQ12162	Factor xa-cleavabl
17	69.6	66.9	198	13	AAQ24986	Sequence encoding
18	69.6	66.9	198	15	AAQ63875	Recombinant hirud
19	68.8	66.2	231	7	AAN60354	Desulphatohirudin
20	66.4	63.8	304	15	AAQ54996	Desulphatohirudin
21	65.8	63.3	567	15	AAQ54997	Leech hirudin mute
22	64.8	62.3	195	16	AAQ81528	Synthetic HVI gene
23	64.8	62.3	257	10	AAN91836	Synthetic hirudin
24	64.8	62.3	257	10	AAN91867	CUP1 promoter, PHO
25	64.8	62.3	1082	15	AAQ64146	Yeast CUP1 promote
26	64.8	62.3	1082	15	AAQ45312	PJDB207/GAPFL-YHIR
27	64.8	62.3	1130	16	AAQ81527	Desulphatohirudin
28	64.6	62.1	219	15	AAQ54995	HV-3 gene. Synthe
29	63	60.6	207	12	AAQ10453	Oleusin-hirudin fu
30	63	60.6	2115	19	AAV44279	Hirudin variant.
31	62.6	60.2	198	15	AAQ63876	DNA encoding the f
32	59.4	57.1	1096	20	AAV83366	Partial Hirudin HV
33	58.6	56.3	183	12	AAQ14927	Synthetic hirudin
34	58.6	56.3	185	12	AAQ14926	Sequence encoding
35	56.4	54.2	195	13	AAQ27889	DNA encoding leech
36	56.4	54.2	198	16	AAQ82762	DNA encoding leech
37	56.4	54.2	204	16	AAQ82764	DNA encoding hirud
38	54.2	52.1	379	6	AAN50395	Sequence encoding
39	54.2	52.1	379	7	AAN60043	Hirudin gene HVI.
40	52.6	50.6	381	6	AAN50396	rHV2-Phe3.Gln33.Ty
41	52.6	50.6	204	14	AAQ38107	Hirudin Phe3 Gln33
42	52.6	50.6	204	16	AAQ83521	Sequence of DNA in
43	52	50.0	526	11	AAQ03237	Encodes hirudin de
44	49	47.1	195	13	AAQ27601	
45	49	47.1	276	13	AAQ27600	

# ALIGNMENTS

RESULT 1	
AAQ12379	AAQ12379 standard; DNA; 726 BP.
ID	AAQ12379; standard; DNA; 726 BP.
XX	
AC	AAQ12379;
XX	
DT	17-SEP-1991 (first entry)
XX	
DE	Hirudin peptide/Protein A fusion gene.
XX	
KW	anticoagulant; fusion protein; ds.
XX	
FH	Key Location/Qualifiers
FT	mat_peptide 1..489
FT	mat_peptide /tag= a
FT	mat_peptide /product= Protein A
FT	mat_peptide 529..720
FT	mat_peptide /tag= b
FT	misc_RNA /product= Hirudin
FT	misc_RNA 490..528
FT	misc_RNA /tag= c
FT	misc_RNA /product= linker oligopeptide y
XX	
PN	DE3942580-A.
XX	
PD	27-JUN-1991.
XX	
PF	22-DEC-1989; 89DE-3942580.
XX	
PR	22-DEC-1989; 89DE-3942580.
XX	
PA	(BADI ) BASF AG.

```

XX PI Korwer W;
XX FT
XX FT WPI; 1991-194236/27.
XX DR P-PSDB; AAR12751.
XX XX
XX XX Hirudin peptide prodn. by cleaving new fusion peptide - of
XX PT hirudin and protein A, expressed in high yield and stable,
XX PT soluble form by transformed E.coli
XX XX
XX XX Example 1; Page 6-7; 9pp; German.
XX CC This sequence is an example of a fusion construct for expression of
XX CC the fusion peptide of the invention. The Met residue in the linker
XX CC oligopeptide allows cleavage by CNBr to release two fragments which
XX CC can be easily separated by IFC affinity chromatography. (The protein
XX CC A component binds to IgG sepharose). Increased yields of hirudin are
XX CC obtained using this fusion construct.
XX XX
XX SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match      92.3%; Score 96; DB 12; Length 726;
Best Local Similarity 100.0%; Pred. No. 3.2e-24;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atgtgtttacactgactgactgaatcccggtcagaacctgtgcctgtgcgaagctct 68
DB 523 atgtgtttacactgactgactgaatcccggtcagaacctgtgcctgtgcgaagctct 582
QY 69 aacgtttgcccaggggcaacaaatgcattcctgggc 104
DB 583 aacgtttgcccaggggcaacaaatgcattcctgggc 618

RESULT 2
AAQ93244
ID AAQ93244 standard; DNA: 468 BP.
AC AAQ93244;
XX
XX 01-NOV-1995 (first entry)
XX Fusion construct of glucoamylase-hirudin DNA.
XX DE
XX DE expression cassette; recombinant protein; production; Hansenula;
XX KW yeast; leader sequence; adaptor; alpha helix; glucoamylase;
XX KW secretion; processing; thrombin inhibitor; hirudin; ds.
XX OS Synthetic.
XX
XX FH Location/Qualifiers
XX FT primer_bind
XX FT complement (1..33)
XX FT /tag= a
XX FT /note= "primer AAQ85840 binding site"
XX FT CDS
XX FT 13..444
XX FT /tag= b
XX FT /product= glucoamylase-hirudin fusion protein
XX FT primer_bind
XX FT 199..223
XX FT /tag= c
XX FT /note= "primer AAQ85841 binding site"
XX FT primer_bind
XX FT complement (220..267)
XX FT /tag= d
XX FT /note= "primer AAQ85842 binding site"
XX FT misc_feature
XX FT 13..228
XX FT /tag= e
XX FT /note= "encodes amino acids 1-72 of glucoamylase"
XX FT misc_feature
XX FT 241..246
XX FT /tag= f
XX FT /note= "encodes processor signal"
XX FT misc_feature
XX FT 247..444
XX FT /tag= g
XX FT /note= "encodes hirudin-h120"

```

```

FT primer_bind 0..468
FT /tag= h
FT /note= "primer AAQ85843 binding site"
XX
XX DE4329969-A.
XX
XX 09-MAR-1995.
XX
XX 04-SEP-1993; 93DE-4329969.
XX
XX 04-SEP-1993; 93DE-4329969.
XX
XX (BADI ) BASF AG.
XX (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX
XX Bollschweiler C, Janowicz ZA, Piontek M, Schweden J;
XX PI Strassersawm, Weydemann U;
XX
XX WPI: 1995-107810/15.
XX P-PSDB; AAR76951.
XX
XX Recombinant protein prodn. in Hansenula yeast - transformed with
XX expression cassette contg. leader, adaptor, processing signal and
XX gene, provides efficient secretion and correct processing
XX
XX Example 1; Fig 1; 10pp; German.
XX
XX AAQ93244 is a DNA fusion product resulting from ligation of an
XX EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding
XX a leader sequence, an adaptor (see AAR71472, amino acids 23-72 of GAM
XX (glucamylase from Schwanniomycetes occidentalis, plus -His-Pro-Leu-Gln at
XX the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,
XX encoding hirudin. AAQ93244 is an example of an expression vector insert
XX of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;
XX P = processor; Gene = structural gene). The cassettes ensure efficient
XX secretion and correct processing of heterologous structural genes in
XX yeast of the genus Hansenula, and so provides high yields of mature
XX proteins and facilitates subsequent purification.
XX
XX SQ Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match      89.4%; Score 93; DB 16; Length 468;
Best Local Similarity 100.0%; Pred. No. 3.3e-23;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gttgtttacactgactgactgaatcccggtcagaacctgtgcctgtgcgaagctctaac 71
DB 247 gttgtttacactgactgactgaatcccggtcagaacctgtgcctgtgcgaagctctaac 306
QY 72 gtttgcggccaggggcaacaaatgcattcctgggc 104
DB 307 gtttgcggccaggggcaacaaatgcattcctgggc 339

RESULT 3
AAQ60746
ID AAQ60746 standard; DNA: 224 BP.
XX
XX AAQ60746;
XX
XX 01-JAN-1980 (first entry)
XX
XX DNA encoding hirudin-like polypeptide.
XX
XX Hirudin; thrombin-antagonist; anticoagulant; ds.
XX
XX Synthetic.
XX
XX DE3445517-A.
XX
XX 19-JUN-1986.

```



PF 13-DEC-1984; 84DE-3445517.

PR 13-DEC-1984; 84DE-3445517.

XX (GENB-) GEN-BIO-TEC GES GEN.

PA Fortkamp E, Rieger M, Sommer R;

PI WPI; 1986-162802/26.

XX P-PSDB; AAP60827.

DR New DNA sequence coding for new hirudin like polypeptide - useful

XX as thrombin antagonist, e.g. for inhibiting blood coagulation

PT Disclosure; Fig. 1; 25pp; German.

XX The sequence encodes a protein with hirudin-like activity. It is a

XX thrombin-antagonist useful for inhibiting blood coagulation and for

CC treating inflammation and/or oedema. It can be prepared in large

CC amounts, unlike natural hirudin which is available only in tiny

CC amounts from leeches. This polypeptide is produced using

CC recombinant DNA techniques and is expressed in a bacterial

CC transformant e.g. Escherichia coli.

XX Sequence 224 BP; 61 A; 51 C; 60 G; 51 T; 1 other;

SQ Query Match 85.0%; Score 88.4; DB 7; Length 224;

XX Best Local Similarity 93.9%; Pred. No. 1.1e-21;

XX Matches 92; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 6 tttatggtgtttacactgactgactgaatccggtcagaaactgtgctgtgcgaagc 65

DB 4 tctatggtgtttacactgactgactgacccgaatctgtgtcagaactgtgctgtgcgaagc 63

QY 66 tctaacgtttgcccagggcgaacaaatgcctctctggg 103

DB 64 tctaacgtttgcccagggcgaacaaatgcctctctggg 101

RESULT 4

AAN60355

ID AAN60355 standard; DNA; 217 BP.

XX AAN60355;

XX 20-JUN-1991 (first entry)

DE Desulphatohirudin (II).

XX Desulphatohirudin; antibodies; thrombin; ds.

XX Synthetic.

XX Key Location/Qualifiers

XX CDS 10..207

XX /\*tag= a

XX /product= hirudin

XX EP168342-A.

XX 15-JAN-1986.

XX 10-JUN-1985; 85EP-0810268.

XX 14-JUN-1984; 84CH-0288284.

XX (CIBA ) CIBA GEIGY AG.

XX Llersch M, Rink H, Markl W, Grutter MG, Meyhack B;

XX WPI; 1986-015589/03.

DR P-PSDB; AAP60395.

XX

PT DNA sequences coding for hirudin and derivs. - and expression

PT vectors, transformed cells, monoclonal antibodies and hybridomas,

XX useful as thrombin inhibitor.

XX Disclosure; Page 10; 123pp; German.

XX The sequence comprises an EcoRI restriction enzyme site at the

XX 5'-end and a BamHI restriction enzyme site at the 3'-end.

XX The sequence may be introduced into a vector for the transformation

XX of hosts, e.g. E. coli. The hirudin can thus be prepared on a

XX large scale. The product and its derivs. are thrombin inhibitors,

XX useful in anticoagulant therapy, esp. when injected at doses

XX of 0.01-0.05 mg/kg.

XX Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

SQ Query Match 83.5%; Score 86.8; DB 7; Length 217;

XX Best Local Similarity 92.9%; Pred. No. 4e-21;

XX Matches 91; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 tttatggtgtttacactgactgactgaatccggtcagaaactgtgctgtgcgaagc 65

DB 7 tctatggtgtttacactgactgactgaatctgtgtcagaactgtgctgtgcgaagc 66

QY 66 tctaacgtttgcccagggcgaacaaatgcctctctggg 103

DB 67 tctaacgtttgcccagggcgaacaaatgcctctctggg 104

RESULT 5

AAN70323

ID AAN70323 standard; DNA; 217 BP.

XX AAN70323;

XX 02-APR-1991 (first entry)

XX Sequence of the desulphatohirudin gene.

XX Anticoagulant; thrombin inhibitor; ds.

XX EP225633-A.

XX 16-JUN-1987.

XX 09-DEC-1986; 86EP-0117098.

XX 29-MAY-1986; 86GB-0013088.

XX 12-DEC-1985; 85GB-0030631.

XX (CIBA ) CIBA GEIGY AG.

XX (PLAN-) PLANTORGAN WERK HEINRICH.

XX (CHRI-) PLANTORGANW CHRISTENSEN.

XX Meyhack B, Markl W, Heim J;

XX WPI; 1987-164868/24.

XX New DNA constructs and hybrid vectors for transformation of yeast

XX etc. - useful for prodn. and secretion of protein with hirudin

XX activity for use as thrombin inhibitors.

XX Example; p44; 146pp; English.

XX The preferred DNA construct of the invention contains the PHO5

XX promoter and a DNA segment consisting of the PHO5 signal sequence

XX upstream of and in reading frame with a DNA sequence coding for

XX mature desulphatohirudin. The segment is under the transcriptional

XX control of the PHO5 promoter and the 3' flanking sequence of the

XX PHO5 gene.



AN81294  
 ID AAN81294 standard; DNA; 229 BP.  
 AC AAN81294;  
 XX  
 DT 17-DEC-1990 (first entry)  
 XX  
 DE Sequence encoding hirudin HV1.  
 XX  
 KW Anticoagulant; thrombin inhibitor; yeast expression cassette; ds.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..4  
 FT /tag= a  
 FT /note="Sticky end"  
 FT misc\_feature complement (226..229)  
 FT /tag= b  
 FT /note="Sticky end"  
 XX  
 PN EP252854-A.  
 PD 13-JAN-1988.  
 XX  
 PF 10-JUL-1987; 87EP-0401649.  
 XX  
 PR 01-DEC-1986; 86FR-0016722.  
 PR 01-JUL-1986; 86FR-0010090.  
 XX  
 PA (TRAN-) TRANSGENE SA.  
 PI Labat N, Loison G, Balland A, Lemoine Y;  
 DR WPI: 1988-008792/02.  
 XX  
 PT New DNA block providing expression of hirudin in yeast -  
 PT comprises hirudin gene, signal and leader sequences and sequence  
 PT for unique proteolytic cleavage site  
 PS Disclosure; ; p; French.  
 XX  
 CC The patent is for a new DNA block providing expression of hirudin (H) in  
 CC yeast. Also new are plasmids contg. the block plus at least one yeast  
 CC origin of replication and yeasts, esp. Saccharomyces cerevisiae,  
 CC transformed with such plasmids.  
 XX  
 SQ Sequence 229 BP; 69 A; 51 C; 57 G; 52 T; 0 other;  
 Query Match 77.7%; Score 80.8; DB 9; Length 229;  
 Best Local Similarity 92.4%; Pred. No. 5.5e-19;  
 Matches 85; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 12 gttgtttacactgactgactgaatccggtcagaaacctgtgctgtgcgaaggctctaac 71  
 DB 20 gtcgtatacactgactgaccggaatccggtcagaaacctgtgctgtgcgaaggctctaac 79.  
 QY 72 gtttgcggccaggggcaaaaatgcctcctggg 103  
 DB 80 gtttgcggccaggggtaacaaatgcctcctggg 111  
 RESULT 9  
 AN81295  
 ID AAN81295 standard; DNA; 238 BP.  
 AC AAN81295;  
 XX  
 DT 17-DEC-1990 (first entry)  
 XX  
 DE Sequence encoding hirudin HV1.  
 XX

KW Anticoagulant; thrombin inhibitor; yeast expression cassette; ss.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..237  
 FT /tag= a  
 XX  
 PN EP252854-A.  
 PD 13-JAN-1988.  
 XX  
 PF 10-JUL-1987; 87EP-0401649.  
 XX  
 PR 01-DEC-1986; 86FR-0016722.  
 PR 01-JUL-1986; 86FR-0010090.  
 XX  
 PA (TRAN-) TRANSGENE SA.  
 PI Labat N, Loison G, Balland A, Lemoine Y;  
 DR WPI: 1988-008792/02.  
 DR P-PSDB; AAP80995.  
 XX  
 PT New DNA block providing expression of hirudin in yeast -  
 PT comprises hirudin gene, signal and leader sequences and sequence  
 PT for unique proteolytic cleavage site  
 PS Disclosure; ; pp; French.  
 XX  
 CC The patent is for a new DNA block providing expression of hirudin (H) in  
 CC yeast. Also new are plasmids contg. the block plus at least one yeast  
 CC origin of replication and yeasts, esp. Saccharomyces cerevisiae,  
 CC transformed with such plasmids.  
 XX  
 SQ Sequence 238 BP; 79 A; 47 C; 55 G; 57 T; 0 other;  
 Query Match 76.0%; Score 79; DB 9; Length 238;  
 Best Local Similarity 89.5%; Pred. No. 2.4e-18;  
 Matches 85; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 QY 9 atgtgtttacactgactgactgaatccggtcagaaacctgtgctgtgcgaaggctct 68  
 DB 1 atgtgtttacactgactgactgactgaatccggtcagaaacctgtgctgtgcgaaggctct 60  
 QY 69 aacgtttgcggccaggggcaaaaatgcctcctggg 103  
 DB 61 aacgtttgcggccaggggtaacaaatgcctcctggg 95  
 RESULT 10  
 AAF61507  
 ID AAF61507 standard; DNA; 227 BP.  
 XX  
 AC AAF61507;  
 XX  
 DT 18-JUN-2001 (first entry)  
 XX  
 DE S. marcescens hirudin DNA fragment.  
 XX  
 KW Hirudin; outer membrane protein; OprF; lambB; fumarate reductase;  
 KW Leu-hirudin; Leu1-Thr2-63-desulfato-hirudin; antithrombotic; ds.  
 XX  
 OS Serratia marcescens.  
 XX  
 PN DE19944870-A1.  
 XX  
 PD 29-MAR-2001.  
 XX  
 PF 18-SEP-1999; 99DE-1044870.  
 XX  
 PR 18-SEP-1999; 99DE-1044870.



PF	07-DEC-1990;	90WO-GB01911.
XX		
PR	07-DEC-1990;	90WO-GB01911.
PR	07-DEC-1989;	89GB-0027722.
XX		
PA	(BRBI-) BRIT BIO-TECHN LTD.	

Query Match 69.2%; Score 72; DB 12; Length 1467;  
Best Local Similarity 84.4%; Pred. No. 1.3e-15;  
Matches 81; Conservative 0; Mismatches 15; Indels

PR 07-DEC-1990; 90WO-GB01911.  
PR 07-DEC-1989; 89GB-0027722.  
XX  
PA (BRBI-) BRIT BIO-TECHN LTD.

Query Match	58.5%	Score 71.2;	DB 12;	Length 201;
Best Local Similarity	85.9%;	Pred. No. 1.3e-15;		
Matches	79;	Conservative	0;	Mismatches 13; Indels 0; Gaps 0;
QY	12	gtgtttacactgactgcactgaatccgggtccgaacactgtgcctgtcgagggtctaac	71	
Db	1	gtgtttacaccactgtactgaatccggacaaacctgttttgtatgaagggtctaac	60	





; FILING DATE: 04-DEC-1990



; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 726 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-262-384A-9

Query Match 92.3%; Score 96; DB 1; Length 726;  
Best Local Similarity 100.0%; Pred. No. 1.5e-25;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atggttgcacactgactgaatccggtcagaacctgtgctgtgcgaagctct 68  
|||||  
Db 523 ATGTTGTTTACACTGACTGCACTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 582  
|||||  
Qy 69 aacgtttgcgcagggcacaataatgcattcctgggc 104  
|||||  
Db 583 AACGTTTGGCGCGAGGCGCAACAATGCATCTCTGGGC 618  
|||||

RESULT 5  
US-08-262-384A-10/C  
; Sequence 10, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 726 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-262-384A-10

Query Match 92.3%; Score 96; DB 1; Length 726;  
Best Local Similarity 100.0%; Pred. No. 1.5e-25;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atggttgcacactgactgaatccggtcagaacctgtgctgtgcgaagctct 68  
|||||  
Db 204 ATGTTGTTTACACTGACTGCACTGAATCCGGTCAGAACCTGTGCTGTGCGAAGGCTCT 145  
|||||

Qy 69 aacgtttgcgcagggcacaataatgcattcctgggc 104  
|||||  
Db 144 AACGTTTGGCGCGAGGCGCAACAATGCATCTCTGGGC 109  
|||||

RESULT 6  
5422249-6  
; Patent No. 5422249  
; APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,  
; MARKUS G.; MEYHACK, BERND  
; TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN  
; INHIBITORS  
; NUMBER OF SEQUENCES: 15  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/121,974  
; FILING DATE: 15-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 700,9978  
; FILING DATE: 10-MAY-1991  
; APPLICATION NUMBER: 582,816  
; FILING DATE: 13-SEP-1990  
; APPLICATION NUMBER: 211,065  
; FILING DATE: 20-JUN-1988  
; APPLICATION NUMBER: 744,453  
; FILING DATE: 13-JUN-1985  
; SEQ ID NO: 6:  
; LENGTH: 109  
5422249-6

Query Match 83.5%; Score 86.8; DB 6; Length 109;  
Best Local Similarity 92.9%; Pred. No. 1.7e-22;  
Matches 91; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 6 ttatggttgtttacactgactgactgaatccggtcagaacctgtgctgtgcgaaggc 65  
|||||  
Db 7 ttcactggtttacacacgactgacgcaatctggtcagaacctgtgctgtgcgaaggt 66  
|||||  
Qy 66 tctaacgtttgcgcagggcacaataatgcattcctggg 103  
|||||  
Db 67 tctaacgtttgcgttcagggttaacaaaatgcattcctggg 104  
|||||

RESULT 7  
5422249-13  
; Patent No. 5422249  
; APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,  
; MARKUS G.; MEYHACK, BERND  
; TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN  
; INHIBITORS  
; NUMBER OF SEQUENCES: 15  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/121,974  
; FILING DATE: 15-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 700,9978  
; FILING DATE: 10-MAY-1991  
; APPLICATION NUMBER: 582,816  
; FILING DATE: 13-SEP-1990  
; APPLICATION NUMBER: 211,065  
; FILING DATE: 20-JUN-1988  
; APPLICATION NUMBER: 744,453  
; FILING DATE: 13-JUN-1985  
; SEQ ID NO: 13:  
; LENGTH: 217  
5422249-13

Query Match 83.5%; Score 86.8; DB 6; Length 217;  
Best Local Similarity 92.9%; Pred. No. 2.1e-22;  
Matches 91; Conservative 0; Mismatches 7; Indels 0; Gaps 0;



;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/EP90/02084  
;; FILING DATE: 04-DEC-1990  
;; APPLICATION NUMBER: US 07861820  
;; FILING DATE: 18-JUN-1992  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 106 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-262-384A-6

Query Match 77.9%; Score 81; DB 1; Length 106;  
Best Local Similarity 100.0%; Pred. No. 2e-20;

Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atggtttttacactgactgaatccggtcagaacctgtgctgtggaagctct 68  
|||  
Db 81 ATGGTTGTTTACACTGACTGCTGAATCCGGTCAGACCTGTGCTGTGCGAAGCTCT 22  
|||  
Qy 69 aacgtttgcggcagggaac 89  
|||  
Db 21 AACGTTTGGCGCAGGGCAAC 1  
|||

## RESULT 11

US-07-859-453E-3  
; Sequence 3, Application US/07859453E  
; Patent No. 5663141

## GENERAL INFORMATION:

;; APPLICANT: Kurfuerst, Manfred  
;; APPLICANT: Ruebsamen, Klaus  
;; APPLICANT: Schmied, Bernhard  
;; APPLICANT: Koerwer, Wolfgang  
;; APPLICANT: Schweden, Juergen  
;; APPLICANT: Hoefken, Hans Wolfgang  
;; TITLE OF INVENTION: Hrudin/polyalkylene glycol  
;; TITLE OF INVENTION: conjugates  
;; NUMBER OF SEQUENCES: 28  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Keil & Weinkauff  
;; STREET: 1101 Connecticut Avenue  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20036

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
;; COMPUTER: IBM compatible  
;; OPERATING SYSTEM: MS-DOS version 7.0  
;; SOFTWARE: Wordperfect version 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/859,453E  
;; FILING DATE: 29-MAY-1992  
;; CLASSIFICATION: 514  
;; CLASSIFICATION: A 61 K 37/64  
;; CLASSIFICATION: A 61 K 47/48

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/EP90/01998  
;; FILING DATE: 22-NOV-1990  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 93 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: unknown  
;; TOPOLOGY: unknown  
;; US-07-859-453E-3

Query Match 76.0%; Score 79; DB 1; Length 93;  
Best Local Similarity 100.0%; Pred. No. 1e-19;

Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 8 tatgtttttacactgactgaatccggtcagaacctgtgctgtggaagctc 67  
|||  
Db 15 TATGGTTGTTTACACTGACTGCTGAATCCGGTCAGACCTGTGCTGTGCGAAGCTC 74  
|||  
Qy 68 taacgtttgcggcagggc 86  
|||  
Db 75 TAACGTTTGGCGCAGGGC 93  
|||

## RESULT 12

US-08-757-439-1/c  
; Sequence 1, Application US/08757439

## GENERAL INFORMATION:

;; APPLICANT: BADZIONG, Werner  
;; APPLICANT: HABERMANN, Paul  
;; APPLICANT: MOELLER, Joerg  
;; APPLICANT: ARETZ, Werner  
;; TITLE OF INVENTION: PROCESS FOR USING THE YEAST ADH II  
;; TITLE OF INVENTION: PROMOTER SYSTEM FOR THE PRODUCTION OF HETEROLOGOUS  
;; TITLE OF INVENTION: PROTEINS IN HIGH YIELDS  
;; NUMBER OF SEQUENCES: 1  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Foley & Lardner  
;; STREET: 3000 K Street, N.W., Suite 500  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20007-5109

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/757,439  
;; FILING DATE: 27-NOV-1996  
;; CLASSIFICATION: 530

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: DE 19544233.4  
;; FILING DATE: 28-NOV-1995

## ATTORNEY/AGENT INFORMATION:

;; NAME: SANDERCOCK, Colin G.  
;; REGISTRATION NUMBER: 31,298  
;; REFERENCE/DOCKET NUMBER: 18748/303/HOCE  
;; TELEPHONE: (202)672-5300  
;; TELEFAX: (202)672-5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 8491 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-757-439-1

Query Match 75.2%; Score 78.2; DB 2; Length 8491;  
Best Local Similarity 91.2%; Pred. No. 8.8e-19;

Matches 83; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 13 ttgtttacactgactgaatccggtcagaacctgtgctgtggaagctctaacg 72  
|||  
Db 6986 TTACGTATACCTGACTGCTGAATCTGCTCAGAACCTGTGCTGTGCGAAGGATCTAACG 6927  
|||  
Qy 73 ttgtgcggcagggaacaaatgcatcctggg 103  
|||  
Db 6926 TTTGCGCCAGGGTAACAAATGCATCTTGG 6896  
|||

```
RESULT 13
5180668-10
; PATENT NO. 5180668
; APPLICANT: CRAUSE, PETER; HABERMANN, PAUL; TRIPIER, DOMINIQUE
; TITLE OF INVENTION: HIRUDIN DERIVATIVE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,422
; FILING DATE: 10-JAN-1989
; SEQ ID NO:10:
; LENGTH: 212
5180668-10

Query Match      75.0%; Score 78; DB 6; Length 212;
Best Local Similarity 94.2%; Pred. No. 3.1e-19;
Matches 81; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 18 tacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggcttaacgtttgc 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 tatactgactgactgaatctgtgcgaacctgtgcctgtgcgaaggatctaaacgtttgc 71

QY 78 gcccagggttaacaaatgcctctgg 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 gcccagggttaacaaatgcctctgg 97

RESULT 14
US-07-982-064-8
; Sequence 8, Application US/07982064
; Patent No. 5919895
; GENERAL INFORMATION:
; APPLICANT: Schmid, G.; Habermann, P.
; TITLE OF INVENTION: Secretion of Hirudin Derivatives
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Collard, Roe & Galgano, P.C.
; STREET: 1077 No. 5919895therr Boulevard
; CITY: Roslyn
; STATE: New York
; COUNTRY: USA
; ZIP: 11576
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/982,064
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/672,791
; FILING DATE: 21-MAR-1991
; APPLICATION NUMBER: GR 40 09 268.2
; FILING DATE: 22 MAR 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Collard, Allison C.
; REGISTRATION NUMBER: 22,532
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W3
; NAME: Galgano, Thomas M.
; REGISTRATION NUMBER: 27,638
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W3
; NAME: Freedman, Edward R.
; REGISTRATION NUMBER: 26,048
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-365-9802
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 227 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

; TOPOLOGY: linear
; MOLECULE TYPE: other DNA
US-07-982-064-8

Query Match      75.0%; Score 78; DB 2; Length 227;
Best Local Similarity 94.2%; Pred. No. 3.1e-19;
Matches 81; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 18 tacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggcttaacgtttgc 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 TATACTGACTGCACTGAATCTGCTCAGAACCTGTGCTGTGCGAAGCACTACACGTTTCG 63

QY 78 gcccagggttaacaaatgcctctggg 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 GGCAGGGTAACAAATGCATCCTTGG 89

RESULT 15
5164304-3
; Patent No. 5164304
; APPLICANT: Johnson, Paul H.; Lazar, Jerome B.; Sohel, Indira
; Waleh, Nahid S.
; TITLE OF INVENTION: METHOD AND VECTORS FOR STABILIZING HIRUDIN
; AND HUMAN LAMININ B1 EXPRESSION
; NUMBER OF SEQUENCES: 26
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/347,545
; FILING DATE: 04-MAY-1989
; SEQ ID NO:3:
; LENGTH: 227
5164304-3

Query Match      74.4%; Score 77.4; DB 6; Length 227;
Best Local Similarity 88.4%; Pred. No. 5.1e-19;
Matches 84; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 9 atggtgttttacactgactgactgaatccggtcagaacctgtgcctgtgcgaaggctct 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 atggtgtttatatacactgattgcacgaatctgtgcagaacctgtgcctgtgtgaaggttcc 70

QY 69 aacgtttgcccagggttaacaaatgcctctggg 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 aacgtatgcgcccagggttaacaaatgcctctggg 105

Search completed: July 15, 2002, 22:31:47
Job time: 18473 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 16:34:09 ; Search time 6165.88 Seconds  
(without alignments)  
227.653 Million cell updates/sec

Title: US-10-053-641-1  
Perfect score: 104  
Sequence: 1 gatccttattgttctttac.....gcaacaatgcatcctgggc 104

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Watch 0%  
Maximum Watch 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_gss:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	33.2	31.9	535	10	BI213620	BI213620 RE19057.5
C 2	31.8	30.6	204	10	BI242456	BI242456 RE35782.5
C 3	31.8	30.6	207	9	AA698346	AA698346 HL04318.5
C 4	31.8	30.6	239	10	BI170090	BI170090 RE11441.5
C 5	31.8	30.6	244	10	BI577602	BI577602 RE71310.5
C 6	31.8	30.6	251	10	BF491529	BF491529 AT28289.5
C 7	31.8	30.6	264	10	BI609601	BI609601 RH14405.5
C 8	31.8	30.6	274	10	BI239641	BI239641 RE36185.5
C 9	31.8	30.6	277	10	BI169275	BI169275 RE10403.5
C 10	31.8	30.6	278	10	BI641092	BI641092 SD24239.5
C 11	31.8	30.6	287	10	BI482355	BI482355 RE65038.5
C 12	31.8	30.6	306	10	BI242439	BI242439 RE39764.5
C 13	31.8	30.6	306	10	BI358576	BI358576 RE45394.5
C 14	31.8	30.6	328	10	BI173930	BI173930 RE17078.5
C 15	31.8	30.6	332	9	AA202819	AA202819 LD03151.5
C 16	31.8	30.6	343	10	BI358523	BI358523 RE45309.5
C 17	31.8	30.6	348	10	BI482786	BI482786 RE65544.5

C 18	31.8	30.6	348	10	BI576148	BI576148 RH32653.5
C 19	31.8	30.6	349	9	AA696815	AA696815 GM08452.5
C 20	31.8	30.6	351	9	AI944520	AI944520 BS03E02.Y
C 21	31.8	30.6	366	9	AI546107	AI546107 LD40936.5
C 22	31.8	30.6	377	9	AA699240	AA699240 HL08021.5
C 23	31.8	30.6	378	10	BI165320	BI165320 RE05175.5
C 24	31.8	30.6	378	10	BI375691	BI375691 RE63554.5
C 25	31.8	30.6	385	10	BI238863	BI238863 RE41957.5
C 26	31.8	30.6	386	9	AI062229	AI062229 GH01366.5
C 27	31.8	30.6	390	10	BI572383	BI572383 RM07881.5
C 28	31.8	30.6	391	9	AA695151	AA695151 RM02322.5
C 29	31.8	30.6	395	10	BI356434	BI356434 RE42587.5
C 30	31.8	30.6	404	9	AI294830	AI294830 LP08292.5
C 31	31.8	30.6	406	10	BI237661	BI237661 RE33928.5
C 32	31.8	30.6	408	10	BI486617	BI486617 RE70067.5
C 33	31.8	30.6	411	10	BI164996	BI164996 RE04788.5
C 34	31.8	30.6	413	10	BI483817	BI483817 RE66766.5
C 35	31.8	30.6	414	9	AA567769	AA567769 HL01745.5
C 36	31.8	30.6	422	10	BI237177	BI237177 RE33355.5
C 37	31.8	30.6	424	9	AA979257	AA979257 LD33625.5
C 38	31.8	30.6	424	10	BI641601	BI641601 SD25016.5
C 39	31.8	30.6	425	10	BI162126	BI162126 RE01532.5
C 40	31.8	30.6	427	10	BI581717	BI581717 RH19254.5
C 41	31.8	30.6	435	10	BI242206	BI242206 RE39482.5
C 42	31.8	30.6	438	10	BI620889	BI620889 RH52143.5
C 43	31.8	30.6	442	9	AA201282	AA201282 LD04219.5
C 44	31.8	30.6	443	9	AA696887	AA696887 GM08634.5
C 45	31.8	30.6	446	9	AA538780	AA538780 LD18383.5

ALIGNMENTS

RESULT 1

BI213620/c

LOCUS

DEFINITION

ATPsyn-beta: FBan0011154 'enzyme' -AND- 'enzyme' located on: 4

102F4-102F4; 04/12/2001, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>

hit genomic AE003846: arm:4 1752420.1050888

estimated-cyto:102E2-102F6: 04/12/2001

Plate: RE190 row: E column: 9

High quality sequence stop: 474.

Location/Qualifiers

1..535

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="RE19057"

/clone\_lib="RE Drosophila melanogaster normalized Embryo

```

pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha TONa"
/Note="Organ: embryo: Vector: pFlc1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
129 a 128 c 140 g 138 t
BASE COUNT
ORIGIN

```

[illegible][illegible]

KEYWORDS	EST.	SOURCE	ORGANISM
	fruit fly.		<i>Drosophila melanogaster</i>
			Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
			1 (bases 1 to 204)
REFERENCE			Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C. J., Nunoo, J., Pacble, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin, G. M.

TITLE	BDGP/HHMI RE Drosophila EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: Stapleton, M. BDGP Lawrence Berkeley National Lab One Cyclotron Rd, Berkeley, CA 94720, USA Fax: 510 486 6798 Email: <a href="http://www.fruitfly.org/EST">http://www.fruitfly.org/EST</a> , <a href="mailto:est@fruitfly.berkeley.edu">est@fruitfly.berkeley.edu</a> Plate: RE:397 row: G column: 10 High quality sequence stop: 198.

```

FEATURES
source
    "49" quality sequence scop: 196.
    Location/Qualifiers
      1..204
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone="RE39782"
        /clone_lib="RE Drosophila melanogaster normalized Embryo
        pf1c-1"
        /sex="male and female"
        /dev_stage="0-24 hours mixed stage embryonic"
        /lab_host="DH5-alpha TONa"
        /note="organ: Vector: pf1c1; Site_1: XhoI; Site_2:
        BamHI; Library was kindly generated by Piero Carninci at
        the RIKEN. The library was normalized and excised using
        Cre recombinase. Plasmid cDNA library."
      50 a 44 c 56 q 54 t
BASE COUNT

```

BASE COUNT	50 a	44 c	56 g	54 t
ORIGIN				

	Query Match	30.6%;	Score 31.8;	DB 10;	Length 204;
	Best Local Similarity	67.2%;	Pred. No. 2.7;		
	Matches 45;	Conservative 0;	Mismatches 22;	Indels 0;	Gaps 0;
Qy	24 gactgcaactgaatccggtctcagaacctgtgcctgtgcgaagctctctaaagtttgcggccag	83			
Db	153 GACTGCACCAATTACGGCCACAATCTTTCCATTTCGGCAGCTGCACCCCTTTCAGCCCTT	94			
Qy	84 ggcaaca	90			
Db	93 GGCAGCA	87			

RESULT	3	
LOCUS	AA698346/c	
DEFINITION	AA698346 207 bp mRNA linear EST 19-APR-2001 HL04318.5prime HL Drosophila melanogaster head Bluescript Drosophila melanogaster cDNA clone HL04318 5prime, mRNA sequence.	
ACCESSION	AA698346	
VERSION	AA698346.1	GI:2701275
KEYWORDS	EST.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	

REFERENCE	1 (bases 1 to 207)
AUTHORS	Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G.M.
TITLE	BDGP/HMMI Drosophila EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: Stapleton, M. BDGP

Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
Plate: 43 row: B column: 6  
High quality sequence stop: 161.

```

FEATURES
source
high quality sequence stop: 101.
location/Qualifiers
1..207
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="HL04318"
/clone.lib="HL Drosophila melanogaster head BlueScript"
/sex="male and female"
/dev_stage="adult"
/lab_host="SOUR"
/note="organ: head-brain & sensory organ; Vector:
BlueScript SK; Site_1: EcoRI; Site_2: XhoI; Constructed
using Stratagene ZAP-cDNA Synthesis kit. Oligo dt-primed
and directionally cloned at EcoRI and XhoI in Bluescript
SK+/-/-."

```

BASE COUNT	49 a	54 g	56 t
ORIGIN	SA("/*")		
Query Match	30.6%	Score 31.8;	DB 9;
Best Local Similarity	67.2%;		Length 207;
Matches	45;	Pred. No. 2.7;	
Conservative	0;	Mismatches	22;
		Indels	0;
		Gaps	0;

24	gactgactgaatccggtcagaacctgtgcctgtgcgaagctctaacgttttgcggccag	83
Qy		
Db		
142	GACTGCACCAATTAGGGCCACAATCTTTCCATTTGCGCAGCTGCAGCCTTTGCAGCCTT	83
Qy		
84	ggcaaca	90
Db		
82	GGCAGCA	76

	RESULT	EST	
BIL70090/c	4	09-JUL-2001	
LOCUS			
BIL70090		mRNA	linear
		239 bp	

```

DEFINITION RE11441.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE11441 5, mRNA sequence.
ACCESSION B1170090
VERSION B1170090.1 GI:14635897
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 239)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin
G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: RE114 row: D column: 5
High quality sequence stop: 232.
Location/Qualifiers
FEATURES
source
1..239
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RE11441"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/notes="Organ: embryo; Vector: pFlc1; Site:1: XhoI; Site:2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 57 a 55 c 63 g 64 t
ORIGIN
Query Match 30.6%; Score 31.8; DB 10; Length 239;
Best Local Similarity 67.2%; Pred. NO. 2.8;
Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 24 gactgcactgaatccggtcagaaacctgtgctgtgcgaaggctctaaagctttgcggccag 83
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153 GACTGCACCAATACGGCCAACTCTTCCATTTGCGGACGCTGCAGCCTTTGCAGCCTT 94
QY 84 ggaaca 90
Db 93 GGCAGCA 87
RESULT 5
LOCUS B1577602
DEFINITION RE17310.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE17310 5 similar to
Atpsyn-beta: FBan011154 GO: [hydrogen-transporting ATP synthase
(GO:0003936); hydrogen-transporting ATP synthase, catalytic core
CF(1) (GO:0005754); hydrogen-transporting ATP synthase (GO:0003936
)] located on: 4 102F4-102F4.; 05/21/2001, mRNA sequence.
ACCESSION B1577602
VERSION B1577602.1 GI:15469024
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 244)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin
G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: RE713 row: A column: 10
High quality sequence stop: 243.
Location/Qualifiers
FEATURES
source
1..244
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RE71310"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/notes="Organ: embryo; Vector: pFlc1; Site:1: XhoI; Site:2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 57 a 55 c 66 g 66 t
ORIGIN
Query Match 30.6%; Score 31.8; DB 10; Length 244;
Best Local Similarity 67.2%; Pred. NO. 2.8;
Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 24 gactgcactgaatccggtcagaaacctgtgctgtgcgaaggctctaaagctttgcggccag 83
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 GACTGCACCAATACGGCCAACTCTTCCATTTGCGGACGCTGCAGCCTTTGCAGCCTT 98
QY 84 ggaaca 90
Db 97 GGCAGCA 91
RESULT 6
LOCUS BF491529
DEFINITION AT28289.5prime AT Drosophila melanogaster adult testes pOTB7
Drosophila melanogaster cDNA clone AT28289 5 similar to
Atpsyn-beta: FBan011154 'enzyme' -AND- 'enzyme' located on: 4
102F4-102F4.; 04/09/2001, mRNA sequence.
ACCESSION BF491529
VERSION BF491529.2 GI:13695056
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 251)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Agbayan, A., Baxter, E., Berman
B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan
D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J.,
Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E.,
Celniker, S. and Rubin, G.M.
TITLE BDGP/HMI AT Drosophila EST Project

```

RH Drosophila EST Project  
d (2001)  
tapleton, M.  
erkeley National Lab  
ron Rd, Berkeley, CA 94720, USA

\_\_\_\_\_



pFlc-1"	
/sex="male and female"	
/dev_stage="0-24 hours mixed stage embryonic"	
/lab_host="DH5-alpha Tona"	
/note="Organ: embryo; Vector: pFlc1; Site:1: XhoI; Site:2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."	
BASE COUNT	66 a 65 c 75 g 68 t
ORIGIN	
Query Match 30.6%; Score 31.8; DB 10; Length 274;	
Best Local Similarity 67.2%; Pred. No. 2.9;	
Matches 45; Conservative 0; Mismatches 22; Idels 0; Gaps 0;	
QY 24	gaatgcactgaatccggctagaacctgtgctgtgcgaagcgtctaagctttgcggcag 83
Db 153	GACTGCACCAATTACGGCCCAACATCTTCCATTTCGGCAGCTGCAGCCCTTGCAGCCTT 94
QY 84	ggcaaca 90
Db 93	GGCAGCA 87
RESULT 9	
BI169275/c	
LOCUS	
DEFINITION	
BI169275 277 bp mRNA linear EST 09-JUL-2001	
REI10403.5prime RE Drosophila melanogaster normalized Embryo pf1c-1	
Drosophila melanogaster cDNA clone REI10403 5 similar to	
ATPSyn-beta: F84N011154 'enzyme' -AND- 'enzyme' located on: 4	
102F4-102F4: 04/11/2001, mRNA sequence.	
ACCESSION	
BI169275.1 GI:14635082	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
fruit fly.	
Drosophila melanogaster	
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
1 (bases 1 to 277)	
REFERENCE	
AUTHORS	
Stapleton,M., Broksstein,P., Hong,L., Tyler,D., Berman,B., Carlson	
J., Champe,M., Chaves,C., Dorsett,V., Farfan,D., Frise,E., George	
R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,	
Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,	
Phouanavong,S., Wan,X., Yu,C., Lewis,S.E., Celnikier,S. and Rubin	
G.M.	
TITLE	
BDGP/HHMI RE Drosophila EST Project	
JOURNAL	
Unpublished (2001)	
COMMENT	
Contact: Stapleton, M.	
BDGP	
Lawrence Berkeley National Lab	
One Cyclotron Rd, Berkeley, CA 94720, USA	
Fax: 510 486 6798	
Email: <a href="http://www.fruitfly.org/EST">http://www.fruitfly.org/EST</a> , <a href="mailto:est@fruitfly.berkeley.edu">est@fruitfly.berkeley.edu</a>	
Plate: RE.104 row: A column: 3	
High quality sequence stop: 234.	
Location/Qualifiers	
1. .277	
/organism="Drosophila melanogaster"	
/db_xref="taxon:7227"	
/clone="REI10403"	
/clone.lib="RE Drosophila melanogaster normalized Embryo	
pf1c-1"	
/sex="male and female"	
/dev_stage="0-24 hours mixed stage embryonic"	
/lab_host="DH5-alpha Tona"	
/note="Organ: embryo; Vector: pFlc1; Site:1: XhoI; Site:2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."	
BASE COUNT	
66 a 66 c 75 g 70 t	
ORIGIN	

1



```

/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="SOLR"
/note="Organ: embryo; Vector: Bluescript SK; Site_1: EcoRI
; Site_2: XhoI; Constructed using Stratagene ZAP-CDNA
Synthesis kit. Oligo dt-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT      76 a      84 c      89 g      83 t
ORIGIN

```

```

Query Match      30.6%; Score 31.8; DB 9; Length 332;
Best Local Similarity 67.2%; Pred. No. 3;
Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY  24 gactgcactgaatccggtcagaacctgtgcctgtgcgaaggctcttaacgtttgcggccag 83
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  175 GACTGCACCAATTACGGCCACAATCTTTCCATTTCGGCAGCTGCAGCCTTTGCAGCCTT 116
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  84 ggcaaca 90
    ||||| ||
Db  115 GGCAGCA 109

```

Search completed: July 15, 2002, 21:48:47  
 Job time: 18878 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:28:32 ; Search time 2368.24 Seconds  
(without alignments)  
945.487 Million cell updates/sec

Title: US-10-053-641-2

Perfect score: 107

Sequence: 1 ctctagagccagatgatcat.....cagtgttaaacaccataaag 107

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_lo.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

C 1	103	96.3	226	12	SYNHIR	M26762 Synthetic h
C 2	99	92.5	129	6	I41183	I41183 Sequence 3
C 3	99	92.5	235	6	I41187	I41187 Sequence 7
C 4	99	92.5	235	6	I41188	I41188 Sequence 8
C 5	99	92.5	726	6	I41189	I41189 Sequence 9
C 6	99	92.5	726	6	I41190	I41190 Sequence 10
C 7	90.8	84.9	220	12	SYNHIRMED	M14964 Synthetic h
C 8	90.8	84.9	224	6	A00630	A00630 Artificial
C 9	90.8	84.9	224	6	A00631	A00631 Artificial
C 10	90.8	84.9	224	6	A04277	A04277 Artificial
C 11	90.8	84.9	224	6	A04278	A04278 Artificial
C 12	90.8	84.9	224	6	A07345	A07345 Synthetic D
C 13	90.8	84.9	224	6	A07346	A07346 Synthetic D
C 14	89.2	83.4	109	6	A34234	A34234 Synthetic h
C 15	89.2	83.4	109	6	A34235	A34235 Synthetic h
C 16	89.2	83.4	217	6	A03695	A03695 Nucleotide
C 17	89.2	83.4	217	6	A03696	A03696 Nucleotide
C 18	89.2	83.4	217	6	A34238	A34238 Synthetic d
C 19	89.2	83.4	217	6	A34239	A34239 Synthetic d
C 20	89.2	83.4	217	6	E00657	E00657 DNA encodin
C 21	84.8	79.3	195	6	A03693	A03693 Nucleotide
C 22	84.8	79.3	195	6	A03694	A03694 Nucleotide
C 23	84.8	79.3	279	6	A01139	A01139 Fusion DNA
C 24	84.8	79.3	279	6	I26634	I26634 Sequence 3
C 25	84.6	79.1	178	6	A13978	A13978 Synthetic D
C 26	84.6	79.1	180	6	A02448	A02448 Nucleotide
C 27	84.6	79.1	180	6	A02449	A02449 Nucleotide
C 28	84.6	79.1	180	6	A13977	A13977 Synthetic D
C 29	83.6	78.1	238	6	A04616	A04616 HV1 gene. 8
C 30	82.2	76.8	225	6	A04613	A04613 Nucleotide
C 31	82.2	76.8	225	6	A04614	A04614 Nucleotide
C 32	82.2	76.8	226	6	A04615	A04615 Nucleotide
C 33	81	75.7	106	6	I41186	I41186 Sequence 6
C 34	79.6	74.4	8491	6	AR031529	AR031529 Sequence
C 35	79.4	74.2	212	6	A13382	A13382 DNA sequenc
C 36	79.4	74.2	212	6	A13383	A13383 DNA sequenc
C 37	79.4	74.2	212	6	A34619	A34619 Synthetic h
C 38	79.4	74.2	212	6	A34620	A34620 Synthetic h
C 39	79.4	74.2	212	6	E00711	E00711 Synthetic D
C 40	79	73.8	93	6	I63540	I63540 Sequence 3
C 41	78.8	73.6	4477	12	SYNEXVECHD	M88535 Expression
C 42	75.6	70.7	223	6	A19994	A19994 SEQ ID NO:
C 43	75.6	70.7	223	6	A20035	A20035 SEQ ID NO:
C 44	75	70.1	211	6	E03003	E03003 DNA encodin
C 45	73.4	68.6	1467	6	A20030	A20030 SEQ ID NO:

ALIGNMENTS

RESULT 1  
SYNHIR/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

CDS

SYNHIR 226 bp DNA linear SYN 27-APR-1993  
Synthetic hirudin gene, complete cds.

M26762.1 GI:208478  
hirudin; proteinase inhibitor; thrombin inhibitor.

Synthetic DNA.  
artificial sequence.

1 (bases 1 to 226)  
Bergmann, C., Dodt, J., Koehler, S., Fink, E. and Gassen, H.G.

Chemical synthesis and expression of a gene coding for hirudin, the  
thrombin-specific inhibitor from the leech Hirudo medicinalis

Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)

87026239 Location/Qualifiers  
1..226  
/organism="synthetic construct"

/db\_xref="taxon:32630"

22..222  
/note="hirudin"

/codon\_start=1





LOCUS	A07346	224 bp	DNA	linear	PAT 03-AUG-1993
DEFINITION	Synthetic DNA for hirudin-similar protein				
ACCESSION	A07346				
VERSION	A07346.1	GI:411427			
KEYWORDS	.				
SOURCE	synthetic	construct.			
ORGANISM	synthetic	construct			







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:46 ; Search time 758.37 Seconds  
(without alignments)  
242.243 Million cell updates/sec

Title: US-10-053-641-2  
Perfect score: 107  
Sequence: 1 cctagagccagatgcat.....cagtgtaaacaccataaag 107

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2002.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
C 1	99	92.5	726 12	AAQ12379 Hirudin peptide/Pr
C 2	98.4	92.0	468 16	AAQ93244 Fusion construct o
C 3	90.8	84.9	224 7	AAN60746 DNA encoding hirud
C 4	89.2	83.4	217 7	AAN60355 Desulphatohirudin
C 5	89.2	83.4	217 8	AAN70323 Sequence of the de
C 6	84.8	79.3	196 8	AAN70319 Sequence encoding
C 7	84.8	79.3	279 12	AAQ13831 MSP signal peptide
C 8	82.2	76.8	229 9	AAN81294 Sequence encoding
C 9	80.4	75.1	238 9	AAN81295 Sequence encoding

C 10	79.6	74.4	227	22	AAF61507	S. marcescens hiru
C 11	75	70.1	210	12	AAQ10452	HV-1 gene. Synthe
C 12	73.4	68.6	1467	12	AAQ12490	Factor Xa-cleavabl
C 13	72.6	67.9	201	12	AAQ12153	Synthetic hirudin
C 14	72.6	67.9	201	13	AAQ25184	Hirudin HV-1. Syn
C 15	72.6	67.9	420	12	AAQ12155	Factor Xa-cleavabl
C 16	72.6	67.9	1458	12	AAQ12162	Factor Xa-cleavabl
C 17	72	67.3	198	13	AAQ24986	Sequence encoding
C 18	72	67.3	198	15	AAQ63875	Recombinant hirudi
C 19	70	63.4	231	7	AAN03354	Desulphatohirudin
C 20	68.2	63.7	567	15	AAQ54997	Leech hirudin mute
C 21	68	63.6	195	16	AAQ81528	Desulphatohirudin
C 22	67.8	63.4	304	15	AAQ54996	Desulphatohirudin
C 23	67.6	63.2	219	15	AAQ54995	Synthetic HV1 gene
C 24	67.2	62.8	257	10	AAN91836	Synthetic hirudin
C 25	67.2	62.8	257	10	AAN91867	CuPI promoter, PHO
C 26	67.2	62.8	1082	15	AAQ64146	Yeast CUP1 promote
C 27	67.2	62.8	1082	15	AAQ45312	PJDB207/GAPFL-YHTR
C 28	67.2	62.8	1130	16	AAQ81527	Oleusin-hirudin fu
C 29	66.4	62.1	2115	19	AAV44279	HV-3 gene. Synthe
C 30	65.4	61.1	207	12	AAQ10453	Hirudin variant.
C 31	65	60.7	198	15	AAQ63876	Partial Hirudin HV
C 32	61	57.0	183	12	AAQ14927	Synthetic hirudin
C 33	61	57.0	185	12	AAQ14926	DNA encoding the f
C 34	60.2	56.3	1096	20	AAV83966	Sequence encoding
C 35	57.6	53.8	195	13	AAQ27889	DNA encoding leech
C 36	57.6	53.8	198	16	AAQ82762	DNA encoding leech
C 37	57.6	53.8	204	16	AAQ82764	DNA encoding hirud
C 38	57.6	53.8	379	6	AAN50395	Sequence encoding
C 39	57.6	53.8	379	7	AAN60043	Hirudin gene HV1.
C 40	57.6	53.8	381	6	AAN50396	rHV2-Phe3.Gln33.Ty
C 41	55.8	52.1	204	14	AAQ38107	Hirudin Phe3 Gln33
C 42	55.8	52.1	204	16	AAQ83521	Sequence of DNA in
C 43	54.4	50.8	526	11	AAQ03237	Encodes hirudin de
C 44	51.4	48.0	195	13	AAQ27601	Encodes hirudin de
C 45	51.4	48.0	276	13	AAQ27600	

ALIGNMENTS

RESULT 1	
AAQ12379/c	
ID	AAQ12379 standard; DNA: 726 BP.
XX	
AC	AAQ12379;
XX	
DT	17-SEP-1991 (first entry)
XX	
DE	Hirudin peptide/Protein A fusion gene.
XX	
KW	anticoagulant; fusion protein; ds.
XX	
FT	Key Location/Qualifiers
FT	mat_peptide 1..489
FT	/*tag= a
FT	/product= Protein A
FT	529..720
FT	mat_peptide
FT	/*tag= b
FT	/product= Hirudin
FT	490..528
FT	/*tag= c
FT	/product= linker oligopeptide Y
XX	
PN	DE3942580-A.
XX	
PD	27-JUN-1991.
XX	
PF	22-DEC-1989; 89DE-3942580.
XX	
PR	22-DEC-1989; 89DE-3942580.
XX	
PA	(BADI ) BASF AG.

```

XX PI Korwer W;
XX DR WPI; 1991-194236/27.
XX DR P-PSDB; AAR12751.
XX PT Hirudin peptide prodn. by cleaving new fusion peptide - of
XX PT hirudin and protein A, expressed in high yield and stable,
XX PT soluble form by transformed E.coli
XX PS Example 1; Page 6-7; 9pp; German.
XX CC This sequence is an example of a fusion construct for expression of
XX CC the fusion peptide of the invention. The Met residue in the linker
XX CC oligopeptide allows cleavage by CNBR to release two fragments which
XX CC can be easily separated by Igg affinity chromatography. (The protein
XX CC A component binds to Igg sepharose). Increased yields of hirudin are
XX CC obtained using this fusion construct.
XX SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match 92.5%; Score 99; DB 12; Length 726;
Best Local Similarity 100.0%; Pred. No. 1.4e-24;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 agagccagcagtgatttgccttgcctgcccgaacgttagagccttcgcacagcagcag 64
Db 621 AGAGCCAGGATGATTTGTTGCCCTGGCGCAACGTTAGAGCCTTCGCACAGGCACAG 562

Oy 65 gttctaccgcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 103
Db 561 GTTCTGACCGGATTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTC 523

RESULT 2
AAQ93244/C
ID AAQ93244 standard; DNA; 468 BP.
XX AC AAQ93244;
XX DT 01-NOV-1995 (first entry)
XX CC Fusion construct of glucoamylase-hirudin DNA.
XX KW expression cassette; recombinant protein; production; Hansenula;
XX KW yeast; leader sequence; adaptor; alpha helix; glucoamylase;
XX KW secretion; processing; thrombin inhibitor; hirudin; ds.
XX OS Synthetic.
XX PH Key
XX FT primer_bind Location/Qualifiers
XX FT complement (1..33)
XX FT /*tag= a
XX FT /*note= "primer AAQ85840 binding site"
XX FT CDS 13..444
XX FT /*tag= b
XX FT /*product= glucoamylase-hirudin fusion protein
XX FT primer_bind 199..223
XX FT /*tag= c
XX FT /*note= "primer AAQ85841 binding site"
XX FT primer_bind complement (220..267)
XX FT /*tag= d
XX FT /*note= "primer AAQ85842 binding site"
XX FT misc_feature 13..228
XX FT /*tag= e
XX FT /*note= "encodes amino acids 1-72 of glucoamylase"
XX FT misc_feature 241..246
XX FT /*tag= f
XX FT /*note= "encodes processor signal"
XX FT misc_feature 247..444
XX FT /*tag= g
XX FT /*note= "encodes hirudin-h120"

```

```

FT primer_bind 0..468
FT /*tag= h
FT /*note= "primer AAQ85843 binding site"
XX PN DE4329969-A.
XX XX 09-MAR-1995.
XX XX 04-SEP-1993; 93DE-4329969.
XX XX 04-SEP-1993; 93DE-4329969.
XX XX (BADI ) BASF AG.
XX XX (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX XX Bollschweiler C, Janowicz ZA, Piontek M, Schweden J;
XX XX Strasserawm, Weydemann U;
XX DR WPI; 1995-107810/15.
XX DR P-PSDB; AAR76951.
XX CC Recombinant protein prodn. in Hansenula yeast - transformed with
XX CC expression cassette contg. leader, adaptor, processing signal and
XX CC gene, provides efficient secretion and correct processing
XX PS Example 1; Fig 1; 10pp; German.
XX CC AAQ93244 is a DNA fusion product resulting from ligation of an
XX CC EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding
XX CC a leader sequence, an adaptor (see AAR71472), amino acids 23-72 of GAM
XX CC (glucamylase from Schwanniomycetes occidentalis, plus -His-Pro-Leu-Gln at
XX CC the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,
XX CC encoding hirudin. AAQ93244 is an example of an expression vector insert
XX CC of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;
XX CC P = processor; Gene = structural gene). The cassettes ensure efficient
XX CC secretion and correct processing of heterologous structural genes in
XX CC yeast of the genus Hansenula, and so provides high yields of mature
XX CC proteins and facilitates subsequent purification.
XX SQ Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match 92.0%; Score 98.4; DB 16; Length 468;
Best Local Similarity 99.0%; Pred. No. 2e-24;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ctctagagcccagcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 60
Db 346 CTTTAGAGCCCGAGGATGATTTGTTGCCCTGGCGCAACGTTAGAGCCTTCGCACAGC 287

Oy 61 acaggttctgaccggattcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 100
Db 286 ACAGGTTCTGACCCGATTTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTC 247

RESULT 3
AAQ60746/C
ID AAQ60746 standard; DNA; 224 BP.
XX AC AAQ60746;
XX XX 01-JAN-1980 (first entry)
XX DT DNA encoding hirudin-like polypeptide.
XX DE Hirudin; thrombin-antagonist; anticoagulant; ds.
XX KW Synthetic.
XX OS DE3445517-A.
XX PN 19-JUN-1986.
XX PD
XX XX

```















Result No.	Query			DB	ID	Description
	Score	Match	Length			
C 1	99	92.5	129	1	US-08-262-384A-3	Sequence 3, Appli
C 2	99	92.5	235	1	US-08-262-384A-7	Sequence 7, Appli
C 3	99	92.5	235	1	US-08-262-384A-8	Sequence 8, Appli
C 4	99	92.5	726	1	US-08-262-384A-9	Sequence 9, Appli
C 5	99	92.5	726	1	US-08-262-384A-10	Sequence 10, Appli
C 6	89.2	83.4	109	6	5422249-6	Patent No. 5422249
C 7	89.2	83.4	217	6	5422249-13	Patent No. 5422249
C 8	89.2	83.4	218	6	5422249-4	Patent No. 5422249
C 9	84.8	79.3	279	1	US-08-186-222-3	Sequence 3, Appli
C 10	81	75.7	106	1	US-08-262-384A-6	Sequence 6, Appli
C 11	79.6	74.4	8491	2	US-08-757-439-1	Sequence 1, Appli
C 12	79.4	74.2	212	6	5180668-10	Patent No. 5180668
C 13	79.4	74.2	237	2	US-07-982-064-8	Sequence 8, Appli
C 14	79	73.8	93	1	US-07-859-453E-3	Sequence 3, Appli
C 15	78.8	73.6	227	6	5164304-3	Patent No. 5164304
C 16	73.4	68.6	1467	1	US-07-854-596B-46	Sequence 46, Appli
C 17	72.6	67.9	201	1	US-07-854-596B-1	Sequence 1, Appli
C 18	72.6	67.9	223	1	US-07-854-596B-7	Sequence 7, Appli
C 19	72.6	67.9	420	1	US-07-854-596B-8	Sequence 8, Appli
C 20	72.6	67.9	1458	1	US-07-854-596B-42	Sequence 42, Appli
C 21	72	67.3	198	1	US-07-910-528-2	Sequence 2, Appli
C 22	72	67.3	198	1	US-08-348-972-2	Sequence 2, Appli
C 23	72	67.3	198	2	US-08-861-459-2	Sequence 2, Appli
C 24	68.2	63.7	563	1	US-08-044-508B-22	Sequence 22, Appli
C 25	68	63.6	195	1	US-08-406-948A-7	Sequence 7, Appli
C 26	67.8	63.4	195	1	US-08-044-508B-27	Sequence 27, Appli
C 27	67.8	63.4	195	1	US-08-044-508B-38	Sequence 38, Appli







```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-262-384A-6

Query Match 75.7%; Score 81; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.8e-20;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 gttgccctggccgcaaacgttagagccttcgcacagggcacaggttctgacggattcagt 82
Db 1 GTTGCCCTGGCCGCAACGTTAGAGCCTTCGCACAGGCACAGGTTCTGACCGGATTTCAGT 60

Qy 83 gcagtcagtgtaaacacaccat 103
Db 61 GCAGTCAGTGTAACAACCAT 81

RESULT 11
US-08-757-439-1
; Sequence 1, Application US/08757439
; Patent No. 5866371
; GENERAL INFORMATION:
; APPLICANT: BADZIONG, Werner
; APPLICANT: HABERMANN, Paul
; APPLICANT: MOELLER, Joerg
; APPLICANT: ARETZ, Werner
; TITLE OF INVENTION: PROCESS FOR USING THE YEAST ADH II
; TITLE OF INVENTION: PROMOTER SYSTEM FOR THE PRODUCTION OF HETEROLOGOUS
; TITLE OF INVENTION: PROTEINS IN HIGH YIELDS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,439
; FILING DATE: 27-NOV-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19544233.4
; FILING DATE: 28-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: SANDERCOCK, Colin G.
; REGISTRATION NUMBER: 31,298
; REFERENCE/DOCKET NUMBER: 18748/303/HOCE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-262-384A-6

Query Match 74.4%; Score 79.6; DB 2; Length 8491;
Best Local Similarity 90.4%; Pred. No. 3.7e-19;
Matches 85; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 6 gagccaggatgcatcttggccctggccgcaaacgttagagccttcgcacagggcacag 65
Db 6893 GATCCAAGGATGCACTTTGTTACCTGGCCGCAACGTTAGATCTCTTCGCACAGGCACAGG 6952

Qy 66 ttctgaccggattcagtgcaagtcagtgtaaacaa 99
Db 6953 TTCTGACCAGATTCACTGTCAGTCAGTCAGTATACGTAA 6986

RESULT 12
5180668-10/c
; Patent No. 5180668
; APPLICANT: CRAUSE, PETER; HABERMANN, PAUL; TRIPIER, DOMINIQUE
; TITLE OF INVENTION: HIRUDIN DERIVATIVE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,422
; FILING DATE: 10-JAN-1989
; SEQ ID NO: 10;
; LENGTH: 212
; 5180668-10

Query Match 74.2%; Score 79.4; DB 6; Length 212;
Best Local Similarity 93.3%; Pred. No. 1.3e-19;
Matches 83; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 6 gagccaggatgcatcttggccctggccgcaaacgttagagccttcgcacagggcacag 65
Db 100 GATCCAAGGATGCACTTTGTTACCTGGCCGCAACGTTAGATCTCTTCGCACAGGCACAGG 41

Qy 66 ttctgaccggattcagtgcaagtcagtgta 94
Db 40 TTCTGACCAGATTCACTGTCAGTCAGTCAGTATA 12

RESULT 13
US-07-982-064-8/c
; Sequence 8, Application US/07982064
; Patent No. 5918895
; GENERAL INFORMATION:
; APPLICANT: Schmid, G.; Habermann, P.
; TITLE OF INVENTION: Secretion of Hirudin Derivatives
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Collard, Roe & Galgano, P.C.
; STREET: 1077 No. 5918895therrn Boulevard
; CITY: Roslyn
; STATE: New York
; COUNTRY: USA
; ZIP: 11576
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/982,064
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/672,791
; FILING DATE: 21-MAR-1991
; APPLICATION NUMBER: GR 40 09 268.2
; FILING DATE: 22 MAR 1990
; ATTORNEY/AGENT INFORMATION:
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:48:47 ; Search time 6165.88 Seconds  
(without alignments)  
234.220 Million cell updates/sec

Title: US-10-053-641-2

Perfect score: 107

Sequence: 1 ctctagagccaggatgcat.....cagttaaacacaccataaag 107

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_gss:\*

13: em\_gss\_hum:\*

14: em\_gss\_inv:\*

15: em\_gss\_pln:\*

16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33.2	31.0	535	10 BI213620	BI213620 RE19057.5
2	31.8	29.7	204	10 BI242456	BI242456 RE39782.5
3	31.8	29.7	207	9 AA698346	AA698346 HL04318.5
4	31.8	29.7	239	10 BI170090	BI170090 RE11441.5
5	31.8	29.7	244	10 BI577602	BI577602 RE71310.5
6	31.8	29.7	251	10 BF491529	BF491529 AT28289.5
7	31.8	29.7	264	10 BI609601	BI609601 RH14405.5
8	31.8	29.7	274	10 BI239641	BI239641 RE36185.5
9	31.8	29.7	277	10 BI169275	BI169275 RE10403.5
10	31.8	29.7	278	10 BI641092	BI641092 SD24239.5
11	31.8	29.7	287	10 BI482355	BI482355 RE65038.5
12	31.8	29.7	306	10 BI242439	BI242439 RE39764.5
13	31.8	29.7	306	10 BI358576	BI358576 RE45394.5
14	31.8	29.7	328	10 BI173930	BI173930 RE17078.5
15	31.8	29.7	332	9 AA02819	AA02819 LD03151.5
16	31.8	29.7	343	10 BI358523	BI358523 RE45309.5
17	31.8	29.7	348	10 BI482786	BI482786 RE65544.5

18	31.8	29.7	348	10 BI576148	BI576148 RH32653.5
19	31.8	29.7	349	9 AA696815	AA696815 GM08452.5
20	31.8	29.7	351	9 AI944520	AI944520 B90302.Y
21	31.8	29.7	366	9 AI546107	AI546107 LD04096.5
22	31.8	29.7	377	9 AA699240	AA699240 HL08021.5
23	31.8	29.7	378	10 BI165320	BI165320 RE03175.5
24	31.8	29.7	378	10 BI375691	BI375691 RE63354.5
25	31.8	29.7	385	10 BI238863	BI238863 RE41957.5
26	31.8	29.7	386	9 AI062229	AI062229 GH01366.5
27	31.8	29.7	390	10 BI572383	BI572383 RH07881.5
28	31.8	29.7	391	9 AA695151	AA695151 GM02322.5
29	31.8	29.7	395	10 BI356434	BI356434 RE42587.5
30	31.8	29.7	404	9 AI294830	AI294830 LP08292.5
31	31.8	29.7	406	10 BI237661	BI237661 RE33928.5
32	31.8	29.7	408	10 BI486617	BI486617 RE70067.5
33	31.8	29.7	411	10 BI164996	BI164996 RE04788.5
34	31.8	29.7	413	10 BI483817	BI483817 RE66766.5
35	31.8	29.7	414	9 AA567769	AA567769 HL01745.5
36	31.8	29.7	422	10 BI237177	BI237177 RE33355.5
37	31.8	29.7	424	9 AA979257	AA979257 LD33625.5
38	31.8	29.7	424	10 BI641601	BI641601 SD25016.5
39	31.8	29.7	425	10 BI162126	BI162126 RE01532.5
40	31.8	29.7	427	10 BI581717	BI581717 RH19254.5
41	31.8	29.7	435	10 BI242206	BI242206 RE39482.5
42	31.8	29.7	438	10 BI620889	BI620889 RH52143.5
43	31.8	29.7	442	9 AA201282	AA201282 LD04219.5
44	31.8	29.7	443	9 AA696887	AA696887 GM08634.5
45	31.8	29.7	446	9 AA538780	AA538780 LD18383.5

## ALIGNMENTS

### RESULT 1

#### BI213620

#### LOCUS

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### COMMENT

#### FEATURES

#### source

BI213620 535 bp mRNA linear EST 11-JUL-2001  
RE19057.5prime RE Drosophila melanogaster normalized Embryo pFLc-1  
Drosophila melanogaster cDNA clone RE19057 5 similar to  
ATPsyn-beta: FBan0011154 'enzyme' -AND- 'enzyme' located on: 4  
102F4-102F4; 04/12/2001, mRNA sequence.

BI213620 GI:14691344

EST.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 535)

Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson

,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George

,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,

Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,

Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celnikier,S. and Rubin

,G.M.

BDGP/HMI RE Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AE003846: arm:4 l752420.1050888]

estimated-cyto:102E2-102F6: 04/12/2001

Plate: RE190 row: E column: 9

High quality sequence stop: 474.

Location/Qualifiers

1..535

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="RE19057"

/clone\_lib="RE Drosophila melanogaster normalized Embryo





[illegible]



```

RESULT 11
BI482355
LOCUS
DEFINITION
    BI482355
    287 bp mRNA linear EST 28-AUG-2001
    R565038.5prime RE Drosophila melanogaster normalized Embryo pF1c-1
    Drosophila melanogaster cDNA clone R565038 5 similar to
    ATPsyn-beta: FBan001154 GO:[hydrogen-transporting ATP synthase
    CF(1) (GO:0003936); hydrogen-transporting ATP synthase, catalytic core
    (GO:0005754); hydrogen-transporting ATP synthase (GO:0003936
    ]] located on: 4 102F4-102F4:: 05/16/2001, mRNA sequence.
    BI482355
    BI482355.1 GI:15324116
    EST.
    fruit fly.
    ORGANISM
        Drosophila melanogaster
        Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
        Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
        Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
        1 (bases 1 to 287)
    AUTHORS
        Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
        ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
        ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
        Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
        Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
        ,G.M.
    TITLE
        BDGP/HMI RE Drosophila EST Project
    JOURNAL
        Unpublished (2001)
    COMMENT
        BDGP
        Lawrence Berkeley National Lab
        One Cyclotron Rd, Berkeley, CA 94720, USA
        Fax: 510 486 6798
        Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
        Plate: RE.650 row: D column: 2
        High quality sequence stop: 221.
        Location/Qualifiers
            1..287
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="R565038"
            /clone_lib="RE Drosophila melanogaster normalized Embryo
            pF1c-1"
            /sex="male and female"
            /dev_stage="0-24 hours mixed stage embryonic"
            /lab_host="DH5-alpha Tona"
            /note="Organ: embryo; Vector: pF1c1; Site_1: XhoI; Site_2:
            BamHI; Library was kindly generated by Piero Carninci at
            the RIKEN. The library was normalized and excised using
            Cre recombinase. Plasmid cDNA library."
            68 a 69 c 79 g 71 t

BASE COUNT
ORIGIN
    Query Match 29.7%; Score 31.8; DB 10; Length 287;
    Best Local Similarity 67.2%; Pred. No. 3.9;
    Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 22 ttgtgccctggcgcacagcttagagccttcgcacaggcacaggtcttcgacggattcag 81
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 87 TGTGTCGAAGGCTGCAAGGCTGCAGCTGCCGCAATGGAAGATTGTGCCCTAATTGG 146

QY 82 tgcagtc 88
    |||||
Db 147 TGCAGTC 153

RESULT 12
BI242439
LOCUS
DEFINITION
    BI242439
    306 bp mRNA linear EST 12-JUL-2001
    R45394.5prime RE Drosophila melanogaster normalized Embryo pF1c-1
    Drosophila melanogaster cDNA clone R539764 5 similar to
    ATPsyn-beta: FBan001154 'enzyme' -AND- 'enzyme' located on: 4
    102F4-102F4:: 05/12/2001, mRNA sequence.
    BI242439
    BI242439.1 GI:15324116
    EST.
    fruit fly.
    ORGANISM
        Drosophila melanogaster
        Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
        Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
        Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
        1 (bases 1 to 306)
    AUTHORS
        Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
        ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
        ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
        Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
        Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
        ,G.M.
    TITLE
        BDGP/HMI RE Drosophila EST Project
    JOURNAL
        Unpublished (2001)
    COMMENT
        BDGP
        Lawrence Berkeley National Lab
        One Cyclotron Rd, Berkeley, CA 94720, USA
        Fax: 510 486 6798
        Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
        Plate: RE.397 row: F column: 4
        High quality sequence stop: 303.
        Location/Qualifiers
            1..306
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="R539764"
            /clone_lib="RE Drosophila melanogaster normalized Embryo
            pF1c-1"
            /sex="male and female"
            /dev_stage="0-24 hours mixed stage embryonic"
            /lab_host="DH5-alpha Tona"
            /note="Organ: embryo; Vector: pF1c1; Site_1: XhoI; Site_2:
            BamHI; Library was kindly generated by Piero Carninci at
            the RIKEN. The library was normalized and excised using
            Cre recombinase. Plasmid cDNA library."
            72 a 71 c 85 g 78 t

BASE COUNT
ORIGIN
    Query Match 29.7%; Score 31.8; DB 10; Length 306;
    Best Local Similarity 67.2%; Pred. No. 4;
    Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 22 ttgtgccctggcgcacagcttagagccttcgcacaggcacaggtcttcgacggattcag 81
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 91 TGTGTCGAAGGCTGCAAGGCTGCAGCTGCCGCAATGGAAGATTGTGCCCTAATTGG 150

QY 82 tgcagtc 88
    |||||
Db 151 TGCAGTC 157

RESULT 13
BI358576
LOCUS
DEFINITION
    BI358576
    306 bp mRNA linear EST 31-JUL-2001
    R45394.5prime RE Drosophila melanogaster normalized Embryo pF1c-1
    Drosophila melanogaster cDNA clone RE45394 5 similar to
    ATPsyn-beta: FBan001154 'enzyme' -AND- 'enzyme' located on: 4
    102F4-102F4:: 05/13/2001, mRNA sequence.
    BI358576
    BI358576.1 GI:15053030
    EST.
    fruit fly.
    ORGANISM
        Drosophila melanogaster
        Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
        Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
        Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
        1 (bases 1 to 306)
    AUTHORS
        Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
        ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
        ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
        Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
        Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
        ,G.M.
    TITLE
        BDGP/HMI RE Drosophila EST Project
    JOURNAL
        Unpublished (2001)
    COMMENT
        BDGP
        Lawrence Berkeley National Lab
        One Cyclotron Rd, Berkeley, CA 94720, USA
        Fax: 510 486 6798
        Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
        Plate: RE.397 row: F column: 4
        High quality sequence stop: 303.
        Location/Qualifiers
            1..306
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="R539764"
            /clone_lib="RE Drosophila melanogaster normalized Embryo
            pF1c-1"
            /sex="male and female"
            /dev_stage="0-24 hours mixed stage embryonic"
            /lab_host="DH5-alpha Tona"
            /note="Organ: embryo; Vector: pF1c1; Site_1: XhoI; Site_2:
            BamHI; Library was kindly generated by Piero Carninci at
            the RIKEN. The library was normalized and excised using
            Cre recombinase. Plasmid cDNA library."
            72 a 71 c 85 g 78 t

BASE COUNT
ORIGIN
    Query Match 29.7%; Score 31.8; DB 10; Length 306;
    Best Local Similarity 67.2%; Pred. No. 4;
    Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 22 ttgtgccctggcgcacagcttagagccttcgcacaggcacaggtcttcgacggattcag 81
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 91 TGTGTCGAAGGCTGCAAGGCTGCAGCTGCCGCAATGGAAGATTGTGCCCTAATTGG 150

QY 82 tgcagtc 88
    |||||
Db 151 TGCAGTC 157

RESULT 13
BI358576
LOCUS
DEFINITION
    BI358576
    306 bp mRNA linear EST 31-JUL-2001
    R45394.5prime RE Drosophila melanogaster normalized Embryo pF1c-1
    Drosophila melanogaster cDNA clone RE45394 5 similar to
    ATPsyn-beta: FBan001154 'enzyme' -AND- 'enzyme' located on: 4
    102F4-102F4:: 05/13/2001, mRNA sequence.
    BI358576
    BI358576.1 GI:15053030
    EST.
    fruit fly.
    ORGANISM
        Drosophila melanogaster
        Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
        Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
        Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
        1 (bases 1 to 306)
    AUTHORS
        Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
        ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
        ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
        Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
        Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
        ,G.M.
    TITLE
        BDGP/HMI RE Drosophila EST Project
    JOURNAL
        Unpublished (2001)
    COMMENT
        BDGP
        Lawrence Berkeley National Lab
        One Cyclotron Rd, Berkeley, CA 94720, USA
        Fax: 510 486 6798
        Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
        Plate: RE.397 row: F column: 4
        High quality sequence stop: 303.
        Location/Qualifiers
            1..306
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="R539764"
            /clone_lib="RE Drosophila melanogaster normalized Embryo
            pF1c-1"
            /sex="male and female"
            /dev_stage="0-24 hours mixed stage embryonic"
            /lab_host="DH5-alpha Tona"
            /note="Organ: embryo; Vector: pF1c1; Site_1: XhoI; Site_2:
            BamHI; Library was kindly generated by Piero Carninci at
            the RIKEN. The library was normalized and excised using
            Cre recombinase. Plasmid cDNA library."
            72 a 71 c 85 g 78 t

BASE COUNT
ORIGIN
    Query Match 29.7%; Score 31.8; DB 10; Length 306;
    Best Local Similarity 67.2%; Pred. No. 4;
    Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 22 ttgtgccctggcgcacagcttagagccttcgcacaggcacaggtcttcgacggattcag 81
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 91 TGTGTCGAAGGCTGCAAGGCTGCAGCTGCCGCAATGGAAGATTGTGCCCTAATTGG 150

QY 82 tgcagtc 88
    |||||
Db 151 TGCAGTC 157

```



```
/sex="male and female"  
/dev_stage="0 to 24 hours mixed stage embryonic"  
/lab_host="SOLR"  
/note="Organ: embryo; Vector: Bluescript SK; Site_1: EcoRI  
; Site_2: XhoI; Constructed using Stratagene ZAP-CDNA  
Synthesis kit. Oligo dt-primed and directionally cloned at  
EcoRI and XhoI in Bluescript SK(+/-)"  
BASE COUNT      76 a      84 c      89 g      83 t  
ORIGIN
```

```
Query Match      29.7%; Score 31.8; DB 9; Length 332;  
Best Local Similarity 67.2%; Pred. No. 4.1;  
Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
Qy 22 ttgtgacctggccgaacgttagagcttcgacagggcacaggttctgacccgattcag 81  
Db 109 TGCTGCCAAGGCTGCAAGGCTGCAGCTGCCGCAATGGAAGATTGTGCCCGTAATTGG 168  
Qy 82 tgcagtc 88  
Db 169 TGCACTC 175
```

Search completed: July 15, 2002, 21:48:48  
Job time: 18879 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:28:32 ; Search time 2368.24 Seconds  
(without alignments)  
980.832 Million cell updates/sec

Title: US-10-053-641-3

Perfect score: 111

Sequence: 1 tctagaggcgcaaaaaatca.....agaacctgcaataatagggc 111

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
-----				

1	37	33.3	195	6	I25751	125751 Sequence 27
2	37	33.3	215	6	A36136	A36136 Sequence 1
3	37	33.3	215	6	E08760	E08760 DNA with Es
4	35.8	32.3	563	6	I25747	I25747 Sequence 22
5	35.8	32.3	567	6	A36139	A36139 Sequence 4
c 6	35.2	31.7	128	6	I41185	I41185 Sequence 5
7	35.2	31.7	226	12	SYNHIR	M26762 Synthetic h
8	35.2	31.7	235	6	I41187	I41187 Sequence 7
9	35.2	31.7	235	6	I41188	I41188 Sequence 8
10	35.2	31.7	726	6	I41189	I41189 Sequence 9
c 11	35.2	31.7	726	6	I41190	I41190 Sequence 10
12	34	30.6	106	6	I41184	I41184 Sequence 4
13	33.2	29.9	152968	2	AC073669	AC073669 Mus muscu
14	33	29.7	59	6	I63542	I63542 Sequence 5
15	33	29.7	59	6	I63546	I63546 Sequence 9
c 16	32.2	29.0	207374	2	AC105822	AC105822 Rattus no
17	32	28.8	279	6	A01139	A01139 Fuson DNA
18	32	28.8	279	6	I26634	I26634 Sequence 3
c 19	31.6	28.5	146809	2	AC091872	AC091872 Homo sapi
c 20	31.6	28.5	155150	9	AC027332	AC027332 Homo sapi
21	31.4	28.3	59	6	I63548	I63548 Sequence 11
22	31.2	28.1	250	14	AF067248	AF067248 Tomato mo
23	31	27.9	195	6	A43186	A43186 Sequence 7
24	31	27.9	195	6	I62835	I62835 Sequence 7
25	31	27.9	198	12	AF274588	AF274588 Synthetic
26	31	27.9	273	12	AF284216	AF284216 Synthetic
27	31	27.9	2157	8	AB000799	AB000799 Arabidops
c 28	31	27.9	2240	8	AY069917	AY069917 Arabidops
29	31	27.9	23404	2	AC017160	AC017160 Drosophi
c 30	31	27.9	24745	8	AC068073	AC068073 Arabidops
c 31	31	27.9	192366	3	AC008369	AC008369 Drosophi
c 32	31	27.9	258873	3	AE003810	AE003810 Drosophi
c 33	30.8	27.7	178	6	A13978	A13978 Synthetic D
34	30.8	27.7	180	6	A02448	A02448 Nucleotide
c 35	30.8	27.7	180	6	A02449	A02449 Nucleotide
36	30.8	27.7	180	6	A13977	A13977 Synthetic D
37	30.8	27.7	212	6	A13382	A13382 DNA sequenc
c 38	30.8	27.7	212	6	A13383	A13383 DNA sequenc
39	30.8	27.7	212	6	A34619	A34619 Synthetic h
c 40	30.8	27.7	212	6	A34620	A34620 Synthetic h
41	30.8	27.7	212	6	E00711	E00711 Synthetic D
42	30.8	27.7	225	6	A04613	A04613 Nucleotide
c 43	30.8	27.7	225	6	A04614	A04614 Nucleotide
44	30.8	27.7	226	6	A04615	A04615 Nucleotide
45	30.8	27.7	238	6	A04616	A04616 HVI gene. 8

#### ALIGNMENTS

RESULT 1	I25751	Sequence 27 from patent US 5552299.	195 bp	DNA	linear	PAT 07-OCT-1996
LOCUS	I25751	Sequence 27 from patent US 5552299.	195 bp	DNA	linear	PAT 07-OCT-1996
DEFINITION	I25751	Sequence 27 from patent US 5552299.	195 bp	DNA	linear	PAT 07-OCT-1996
ACCESSION	I25751	Sequence 27 from patent US 5552299.	195 bp	DNA	linear	PAT 07-OCT-1996
VERSION	I25751.1	GI:1605621	195 bp	DNA	linear	PAT 07-OCT-1996
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 195)					
AUTHORS	Ott, Ian., Klupp, T., Molnar, Ian., Patthy, Aas., Barta, Ian., Bark, O n ee T oth, Z., Ambrus, Gabor., Sal at, Janos., Tegdes, A., Moravcsik, I., Esgy ud, C., Albrecht, Karnly., K oncz Ol, Kalman., Vincze, A., Barab as, E., M at e, Gorgy., Kiss, Gorgy. B., Kiss, Peter., P Oly, Kalman., Erdei, Janos., Guly as, E. and Zilahi, E. Plasmids and process for producing recombinant desulphatohirudin HV-1 peptides Patent: US 5552299-A 27 03-SEP-1996; Location/Qualifiers 1. .195 /organism="unknown"					
TITLE						
JOURNAL						
FEATURES						
BASE COUNT	58 a	40 c	45 g	51 t	1 others	

\_\_\_\_\_



I41188/c  
LOCUS I41188 235 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 8 from patent US 5624822.  
ACCESSION I41188  
VERSION I41188.1 GI:2081778  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 235)  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 8 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..235  
BASE COUNT 42 a 58 c 64 g 71 t  
ORIGIN

Query Match 31.7%; Score 35.2; DB 6; Length 235;  
Best Local Similarity 83.3%; Pred. No. 0.58;  
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 tctagagcgcaaaatcaatcggttactggcgaagggtactccacaa 48  
||| ||||||||| || ||||||||| ||||||||| || ||  
Db 113 TCTGACGCGCAAAACCAAGTGGTACTGGCGAAGGTACCCCGAAA 66

RESULT 10  
I41189  
LOCUS I41189 726 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 9 from patent US 5624822.  
ACCESSION I41189  
VERSION I41189.1 GI:2081779  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 726)  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 9 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..726  
BASE COUNT 281 a 171 c 126 g 148 t  
ORIGIN

Query Match 31.7%; Score 35.2; DB 6; Length 726;  
Best Local Similarity 83.3%; Pred. No. 0.57;  
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 tctagagcgcaaaatcaatcggttactggcgaagggtactccacaa 48  
||| ||||||||| || ||||||||| ||||||||| || ||  
Db 619 TCTGACGCGCAAAACCAAGTGGTACTGGCGAAGGTACCCCGAAA 666

RESULT 11  
I41190/c  
LOCUS I41190 726 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 10 from patent US 5624822.  
ACCESSION I41190  
VERSION I41190.1 GI:2081780  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 726)  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 10 29-APR-1997;

FEATURES source Location/Qualifiers  
1..726  
BASE COUNT 149 a 126 c 171 g 280 t  
ORIGIN

Query Match 31.7%; Score 35.2; DB 6; Length 726;  
Best Local Similarity 83.3%; Pred. No. 0.57;  
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 tctagagcgcaaaatcaatcggttactggcgaagggtactccacaa 48  
||| ||||||||| || ||||||||| ||||||||| || ||  
Db 108 TCTGACGCGCAAAACCAAGTGGTACTGGCGAAGGTACCCCGAAA 61

RESULT 12  
I41184  
LOCUS I41184 106 bp DNA linear PAT 13-MAY-1997  
DEFINITION Sequence 4 from patent US 5624822.  
ACCESSION I41184  
VERSION I41184.1 GI:2081774  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 106)  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 4 29-APR-1997;  
FEATURES Location/Qualifiers  
source 1..106  
BASE COUNT 36 a 28 c 28 g 14 t  
ORIGIN

Query Match 30.6%; Score 34; DB 6; Length 106;  
Best Local Similarity 88.1%; Pred. No. 1.5;  
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 7 ggcgaatcaatcggttactggcgaagggtactccacaa 48  
||| ||||||||| || ||||||||| ||||||||| || ||  
Db 4 GCGCAAAACCAAGTGGTACTGGCGAAGGTACCCCGAAA 45

RESULT 13  
AC073669  
LOCUS AC073669 152968 bp DNA linear HTG 29-JUN-2000  
DEFINITION Mus musculus clone CT7-273D11, WORKING DRAFT SEQUENCE, 39 unordered pieces.  
ACCESSION AC073669  
VERSION AC073669.1 GI:8810286  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
AUTHORS 1 (bases 1 to 152968)  
TITLE Sequencing of Mouse  
JOURNAL DOE Joint Genome Institute.  
REFERENCE 2 (bases 1 to 152968)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information

Center Project Name: 1058096  
Center clone name: RG-MBAC\_273D11

-----  
Summary Statistics

Consensus quality: 118135 bases at least Q40  
Consensus quality: 134236 bases at least Q30  
Consensus quality: 137982 bases at least Q20  
Estimated insert size: 100000; pulse field gel estimation  
Estimated insert size: 149168; sum-of-contigs estimation  
Quality coverage: 7.41 in Q20 bases; pulse field gel estimation  
Quality coverage: 4.96 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 39 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1093: contig of 1093 bp in length  
\* 1094 1193: gap of unknown length  
\* 1194 2195: contig of 1002 bp in length  
\* 2196 2295: gap of unknown length  
\* 2296 3335: contig of 1040 bp in length  
\* 3336 3435: gap of unknown length  
\* 3436 4873: contig of 1438 bp in length  
\* 4874 4974: gap of unknown length  
\* 4975 6084: contig of 1111 bp in length  
\* 6085 6184: gap of unknown length  
\* 6185 7742: contig of 1558 bp in length  
\* 7743 7842: gap of unknown length  
\* 7843 8945: contig of 1103 bp in length  
\* 8946 9046: gap of unknown length  
\* 9047 10120: contig of 1075 bp in length  
\* 10121 10221: gap of unknown length  
\* 10222 11475: contig of 1255 bp in length  
\* 11476 11575: gap of unknown length  
\* 11576 12699: contig of 1124 bp in length  
\* 12700 12799: gap of unknown length  
\* 12800 13987: contig of 1188 bp in length  
\* 13988 14087: gap of unknown length  
\* 14088 15182: contig of 1095 bp in length  
\* 15183 15282: gap of unknown length  
\* 15283 16467: contig of 1185 bp in length  
\* 16468 16567: gap of unknown length  
\* 16568 18435: contig of 1868 bp in length  
\* 18436 18535: gap of unknown length  
\* 18536 20772: contig of 2237 bp in length  
\* 20773 20872: gap of unknown length  
\* 20873 22113: contig of 1341 bp in length  
\* 22114 22313: gap of unknown length  
\* 22314 23495: contig of 1182 bp in length  
\* 23496 23595: gap of unknown length  
\* 23596 25217: contig of 1622 bp in length  
\* 25218 25317: gap of unknown length  
\* 25318 27409: contig of 2092 bp in length  
\* 27410 27509: gap of unknown length  
\* 27510 28855: contig of 1356 bp in length  
\* 28856 28965: gap of unknown length  
\* 28966 31579: contig of 2614 bp in length  
\* 31580 31679: gap of unknown length  
\* 31680 34419: contig of 2740 bp in length  
\* 34420 34519: gap of unknown length  
\* 34520 38002: contig of 3483 bp in length  
\* 38003 38102: gap of unknown length  
\* 38103 42293: contig of 4191 bp in length  
\* 42294 42393: gap of unknown length  
\* 42394 47356: contig of 4963 bp in length  
\* 47357 47456: gap of unknown length  
\* 47457 51699: contig of 4243 bp in length  
\* 51700 51799: gap of unknown length  
\* 51800 55044: contig of 3245 bp in length  
\* 55045 55144: gap of unknown length

\* 55145 59713: contig of 4569 bp in length  
\* 59714 59813: gap of unknown length  
\* 59814 64897: contig of 5084 bp in length  
\* 64898 64997: gap of unknown length  
\* 64998 69683: contig of 4686 bp in length  
\* 69684 69783: gap of unknown length  
\* 69784 74689: contig of 4906 bp in length  
\* 74690 74789: gap of unknown length  
\* 74790 80906: contig of 6117 bp in length  
\* 80907 81006: gap of unknown length  
\* 81007 90922: contig of 9916 bp in length  
\* 91022 91023: gap of unknown length  
\* 91024 100745: contig of 9723 bp in length  
\* 100746 100845: gap of unknown length  
\* 100846 107914: contig of 7069 bp in length  
\* 107915 108014: gap of unknown length  
\* 108015 117904: contig of 9890 bp in length  
\* 117905 118004: gap of unknown length  
\* 118005 131952: contig of 13948 bp in length  
\* 131953 132052: gap of unknown length  
\* 132053 140157: contig of 8105 bp in length  
\* 140158 140257: gap of unknown length  
\* 140258 152968: contig of 12711 bp in length.

## FEATURES

## source

1. 152968  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="Ct7-273D11"  
/clone\_11b="CtBCJ7 mouse BAC library"  
BASE COUNT 39470 a 32372 c 31940 g 45341 t 3845 others  
ORIGIN

Query Match 29.9%; Score 33.2; DB 2; Length 152968;

Best Local Similarity 59.6%; Pred. No. 2.3; Mismatches 0; Gaps 0; Indels 0; Gaps 0;

Matches 56; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 4 agagcgcaaaaaatcaatgcgttactggcgaaggtactccacaggtttgaatcccg 63

Db 38124 ATACACTAAAGACACATCCAGCCAGGCTACTATACCTGCCAAACTCTCAATTACCA 38183

Qy 64 aagaatagaacccgcagtcctccgcagcagagaacc 97

Db 38184 TAGAGGAGAACCCAAAGTATTCCAGCAGCAAAACC 38217

## RESULT 14

LOCUS I63542 I63542 Sequence 5 from patent US 5663141. linear PAT 07-OCT-1997

DEFINITION I63542

ACCESSION I63542

VERSION I63542.1 GI:2481115

KEYWORDS

SOURCE Unknown.

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 59)

AUTHORS Kurfuerst, M., Ruebaamen, K., Schmied, B., Koerwer, W., Schweden, J. and

Hoeffken, H. Wolfgang.

TITLE Hirusin/polyalkylene glycol conjugates and hirusin mutants

JOURNAL Patent: US 5663141-A 5 02-SEP-1997;

FEATURES

Location/Qualifiers

1. 59

source

/organism="unknown"

BASE COUNT 19 a 14 c 16 g 10 t

## ORIGIN

Query Match 29.7%; Score 33; DB 6; Length 59;

Best Local Similarity 87.8%; Pred. No. 3.2; Mismatches 0; Gaps 0; Indels 0; Gaps 0;

Matches 36; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tctagagcgcaaaaaatcaatgcgttactggcgaaggtac 41

Search completed: July 15, 2002, 22:28:44  
Job time: 18970 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:47 ; Search time 758.37 Seconds  
(without alignments)  
251.299 Million cell updates/sec

Title: US-10-053-641-3

Perfect score: 111

Sequence: 1 tctaggaggcaaaaaatac.....agaacctgcaataataggcc 111

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	ID	Description
1	37	33.3	219 15	AAQ54995
2	35.8	32.3	567 15	AAQ54997
3	35.2	31.7	726 12	AAQ12379
4	33.2	29.9	468 16	AAQ93244
5	32	28.8	279 12	AAQ13831
6	31	27.9	195 16	AAQ81528
7	31	27.9	1984 21	AAZ98321
8	31	27.9	2157 21	AAZ51103
9	30.8	27.7	227 22	AAAF61507

10	30.8	27.7	229	9	AAAN81294	Sequence encoding
11	30.8	27.7	238	9	AAAN81295	Sequence encoding
12	30.6	27.6	231	7	AAAN60354	Desulphatohirudin
13	30.4	27.4	195	13	AAQ27601	Encodes hirudin de
14	30.4	27.4	196	8	AAAN70319	Sequence encoding
15	30.4	27.4	208	10	AAAN91851	Synthetic DNA enco
16	30.4	27.4	217	7	AAAN60355	Desulphatohirudin
17	30.4	27.4	217	8	AAAN70323	Sequence of the de
18	30.4	27.4	224	7	AAAN60746	DNA encoding hirud
19	30.4	27.4	276	13	AAQ27600	Encodes hirudin de
20	30.4	27.4	525	11	AAQ06873	Secretion plasmid
21	30.4	27.4	526	11	AAQ03237	Sequence of DNA in
22	30	27.0	358	6	AAAN50397	Synthetic hirudin
23	29.8	26.8	43	14	AAQ43368	Hirudin oligonucle
24	29.4	26.5	257	10	AAAN91836	Synthetic hirudin
25	29.4	26.5	1082	15	AAQ64146	CUP1 promoter, PHO
26	29.4	26.5	1082	15	AAQ45312	Yeast CUP1 promote
27	29.4	26.5	1130	16	AAQ81527	PJDB207/GAPFL-YHIR
28	29.2	26.3	24158	21	AAAN1532	N. meningitidis pa
29	29.2	26.3	349980	21	AAAF21609	Neisseria meningit
30	28.8	25.9	198	13	AAQ24986	Sequence encoding
31	28.8	25.9	198	13	AAQ63876	Hirudin variant.
32	28.8	25.9	210	12	AAQ10452	HV-1 gene. Synthe
33	28.2	25.4	201	12	AAQ12153	Synthetic hirudin
34	28.2	25.4	201	13	AAQ25184	Hirudin HV-1. Syn
35	28.2	25.4	420	12	AAQ12155	Factor Xa-cleavabl
36	28.2	25.4	1458	12	AAQ12162	Factor Xa-cleavabl
37	28.2	25.4	1467	12	AAQ12490	Factor Xa-cleavabl
38	28	25.2	183	12	AAQ14927	Partial Hirudin HV
39	28	25.2	7157	22	ABAI9506	Human nervous syst
40	27.8	25.0	944	21	AAAC51426	Arabidopsis thalia
41	27.8	25.0	1335	21	AAAC36967	Arabidopsis thalia
42	27.8	25.0	422	21	AAAC38173	Arabidopsis thalia
43	27.4	24.7	959	21	AAAC33229	Arabidopsis thalia
44	27.4	24.7	8900	22	AAAL36092	Human musculoskele

#### ALIGNMENTS

RESULT 1

AAQ54995

ID AAQ54995 standard; DNA; 219 BP.

XX AAQ54995;

AC AAQ54995;

XX 11-JUL-1994 (first entry)

DT Desulphatohirudin gene with E. coli-Saccharomyces codon usage.

DE Hirudin; HV-1; E. coli; codon usage: desulphatohirudin; HV01; 33asp;

XX 33 Asp; biological activity; thrombosis; thromboembolism; ds.

OS Hirudo medicinalis.

XX Key Location/Qualifiers

EH CDS 14..208

FT /\*tag= a

FT /product= Hirudin\_HV-1

XX EP576792-A.

XX 05-JAN-1994.

PD 13-APR-1993; 93EP-0105848.

XX 09-APR-1992; 92HU-0001200.

XX (BIOG ) BIOGAL GYOGYSZERGYAR.

PA Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;

XX Botond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoal K;

```

PI Mate G, Molnari, Moravcsik I, Ott I, Patthy A;
PI Poelya K, Salat J, Tegdes A, Vincze A, Zilahi E;
XX
XX WPI: 1994-009153/02.
XX P-PSDB; AAR47488.
XX
XX Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using
XX E. coli, Saccharomyces and Streptomyces hosts, for increased
XX yields
XX
XX Disclosure: Page 44; 79pp; English.
XX
XX This sequence represents the hirudin HV-1 gene designed on the
XX basis on E. coli codon usage. This sequence was used in the
XX production of desulphatohirudin HV01 33asp and desulphatohirudin HV-1
XX 33 Asn. The expressed peptides produced using this sequence have the
XX same biological activity as natural hirudin and can be used in the
XX treatment of thrombosis, thromboembolism, etc. Using naturalised
XX sequences such as this, large amounts of hirudin may be produced
XX stably, with the highest production level achieved being 140-180
XX mg/litre of culture.
XX
XX Sequence 219 BP; 63 A; 46 C; 51 G; 58 T; 1 other;
XX
XX Query Match 33.3%; Score 37; DB 15; Length 219;
XX Best Local Similarity 85.1%; Pred. No. 0.0019;
XX Matches 40; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
Oy 1 tctagagcgcaaaatcaatgcgttactggtcgcgaaggtactccaca 47
Db 107 tctacggtgaaaaaatcaatgtctcactggtcgcgaaggtactccaaa 153
RESULT 2
AAQ54997
ID AAQ54997 standard; DNA; 567 BP.
XX
XX AAQ54997;
XX
XX 11-JUL-1994 (first entry)
XX
XX Desulphatohirudin gene expression/excretion cassette.
XX
XX Hirudin; HV-1; E. coli; codon usage; desulphatohirudin; HV01; 33asp;
XX 33 Asn; biological activity; thrombosis; thromboembolism; ss.
XX
XX Hirudo medicinalis.
XX
XX Key Location/Qualifiers
XX -35_signal 43..48
XX /*tag= a
XX -10_signal 68..73
XX /*tag= b
XX misc_signal 83
XX /*tag= c
XX /note= "Transcription initiation site"
XX RBS 260..266
XX /*tag= d
XX /note= "Shine-Dalgarno sequence"
XX sig_peptide 269..352
XX /*tag= e
XX mat_peptide 353..550
XX /*tag= f
XX /product= Hirudin
XX
XX EP576792-A.
XX
XX 05-JAN-1994.
XX
XX 13-APR-1993; 93EP-0105848.
XX
XX 09-APR-1992; 92HU-0001200.
XX

```

```

XX
XX (BIOG ) BIOGAL GYOGYSZERGIAR.
XX
XX Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;
XX Botond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;
XX Mate G, Molnari, Moravcsik I, Ott I, Patthy A;
XX Poelya K, Salat J, Tegdes A, Vincze A, Zilahi E;
XX
XX WPI: 1994-009153/02.
XX P-PSDB; AAR47490.
XX
XX Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using
XX E. coli, Saccharomyces and Streptomyces hosts, for increased
XX yields
XX
XX Disclosure: Page 48; 79pp; English.
XX
XX This sequence represents the an expression/secretion cassette for
XX the expression of hirudin HV-1. This sequence was used in the
XX production of desulphatohirudin HV01 33asp and desulphatohirudin HV-1
XX 33 Asn. The expressed peptides produced using this sequence have the
XX same biological activity as natural hirudin and can be used in the
XX treatment of thrombosis, thromboembolism, etc. Using naturalised
XX sequences such as this, large amounts of hirudin may be produced
XX stably, with the highest production level achieved being 140-180
XX mg/litre of culture.
XX
XX Sequence 567 BP; 162 A; 123 C; 134 G; 148 T; 0 other;
XX
XX Query Match 32.3%; Score 35.8; DB 15; Length 567;
XX Best Local Similarity 85.1%; Pred. No. 0.0086;
XX Matches 40; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Oy 1 tctagagcgcaaaatcaatgcgttactggtcgcgaaggtactccaca 47
Db 446 tctgacggtgaaaaaatcaatgtctcactggtcgcgaaggtactccaaa 492
RESULT 3
AAQ12379
ID AAQ12379 standard; DNA; 726 BP.
XX
XX AAQ12379;
XX
XX 17-SEP-1991 (first entry)
XX
XX Hirudin peptide/Protein A fusion gene.
XX
XX anticoagulant; fusion protein; ds.
XX
XX Key Location/Qualifiers
XX mat_peptide 1..489
XX /*tag= a
XX /product= Protein A
XX mat_peptide 529..720
XX /*tag= b
XX /product= Hirudin
XX misc_rna 490..528
XX /*tag= c
XX /product= linker oligopeptide Y
XX
XX DE3942580-A.
XX
XX 27-JUN-1991.
XX
XX 22-DEC-1989; 89DE-3942580.
XX
XX 22-DEC-1989; 89DE-3942580.
XX
XX (BADI ) BASF AG.
XX
XX Korwer W;
XX

```









Qy 7 ggcgaataatcaatgcgttactggcgaagggtactccacaa 48  
 || ||||| || || ||||| ||||| || || ||  
 Db 100 ggtgaaaaaacccagtcggttactggcgaagggtactcccgaaa 141

RESULT 10  
 AAN81294  
 ID AAN81294 standard; DNA; 229 BP.  
 XX  
 AC AAN81294;  
 XX  
 DT 17-DEC-1990 (first entry)  
 XX  
 DE Sequence encoding hirudin HV1.  
 XX  
 KW Anticoagulant; thrombin inhibitor; yeast expression cassette; ds.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..4  
 FT /\*tag= a  
 FT /note="Sticky end"  
 FT complement (226..229)  
 FT /\*tag= b  
 FT /note="Sticky end"  
 XX  
 EP252854-A.  
 PN  
 XX  
 PD 13-JAN-1988.  
 XX  
 PF 10-JUL-1987; 87EP-0401649.  
 XX  
 PR 01-DEC-1986; 86FR-0016722.  
 PR 01-JUL-1986; 86FR-0010090.  
 XX  
 (TRAN-) TRANSGENE SA.  
 PA  
 XX  
 PI Labat N, Loison G, Balland A, Lemoine Y;  
 XX  
 DR WPI; 1988-008792/02.  
 XX  
 PT New DNA block providing expression of hirudin in yeast -  
 PT comprises hirudin gene, signal and leader sequences and sequence  
 PT for unique proteolytic cleavage site  
 XX  
 PS Disclosure; ; p; French.  
 XX  
 CC The patent is for a new DNA block providing expression of hirudin (H) in  
 CC yeast. Also new are plasmids contg. the block plus at least one yeast  
 CC origin of replication and yeasts, esp. Saccharomyces cerevisiae,  
 CC transformed with such plasmids.  
 XX  
 SQ Sequence 229 BP; 69 A; 51 C; 57 G; 52 T; 0 other;

Query Match 27.7%; Score 30.8; DB 9; Length 229;  
 Best Local Similarity 83.3%; Pred. No. 0.3;  
 Matches 35; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 Qy 7 ggcgaataatcaatgcgttactggcgaagggtactccacaa 48  
 || ||||| || || ||||| ||||| || || ||  
 Db 119 ggtgaaaaaacccagtcggttactggcgaagggtactcccgaaa 160

RESULT 11  
 AAN81295  
 ID AAN81295 standard; DNA; 238 BP.  
 XX  
 AC AAN81295;  
 XX  
 DT 17-DEC-1990 (first entry)

XX  
 DE Sequence encoding hirudin HV1.  
 XX  
 KW Anticoagulant; thrombin inhibitor; yeast expression cassette; ss.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..237  
 FT /\*tag= a  
 XX  
 EP252854-A.  
 PN  
 XX  
 PD 13-JAN-1988.  
 XX  
 PF 10-JUL-1987; 87EP-0401649.  
 XX  
 PR 01-DEC-1986; 86FR-0016722.  
 PR 01-JUL-1986; 86FR-0010090.  
 XX  
 (TRAN-) TRANSGENE SA.  
 PA  
 XX  
 PI Labat N, Loison G, Balland A, Lemoine Y;  
 XX  
 DR WPI; 1988-008792/02.  
 DR P-PSDB; AAP80995.  
 XX  
 PT New DNA block providing expression of hirudin in yeast -  
 PT comprises hirudin gene, signal and leader sequences and sequence  
 PT for unique proteolytic cleavage site  
 XX  
 PS Disclosure; ; pp; French.  
 XX  
 CC The patent is for a new DNA block providing expression of hirudin (H) in  
 CC yeast. Also new are plasmids contg. the block plus at least one yeast  
 CC origin of replication and yeasts, esp. Saccharomyces cerevisiae,  
 CC transformed with such plasmids.  
 XX  
 SQ Sequence 238 BP; 79 A; 47 C; 55 G; 57 T; 0 other;

Query Match 27.7%; Score 30.8; DB 9; Length 238;  
 Best Local Similarity 83.3%; Pred. No. 0.3;  
 Matches 35; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 Qy 7 ggcgaataatcaatgcgttactggcgaagggtactccacaa 48  
 || ||||| || || ||||| ||||| || || ||  
 Db 103 ggtgaaaaaacccagtcggttactggcgaagggtactcccgaaa 144

RESULT 12  
 AAN60354  
 ID AAN60354 standard; DNA; 231 BP.  
 XX  
 AC AAN60354;  
 XX  
 DT 20-JUN-1991 (first entry)  
 XX  
 DE Desulphatohirudin (I).  
 XX  
 KW Desulphatohirudin; antibodies; thrombin; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN EP168342-A.  
 XX  
 PD 15-JAN-1986.  
 XX  
 PF 10-JUN-1985; 85EP-0810268.  
 PR 14-JUN-1984; 84CH-0288284.  
 XX  
 PA (CIBA ) CIBA GEIGY AG.



```

XX Synthetic DNA encoding desulphatohirudin.
DE Desulphatohirudin; synthetic gene; thrombin inhibitor; HV-1 gene;
KW plasmid p3010; plasmid p4014.
XX
XX JP01247092-A.
XX
XX 02-OCT-1989.
XX
XX 29-MAR-1988; 88JP-0073200.
XX
XX 29-MAR-1988; 88JP-0073200.
XX
XX (MITK ) MITSUI TOATSU CHEM. INC.
XX
XX WPI; 1989-330037/45.
XX
XX Prepn. of desulphatohirudin protein used as thrombin inhibitor - by
PT introducing specific DNA into expression vector, inserting obtd.
PT recombinant plasmid into host and incubating.
XX
XX Claim 1; page 485; 6pp; Japanese.
XX
XX The synthetic gene is constructed by dividing the hirudin HV-1 gene into
CC 7 parts and ligating to form the gene. It is inserted into pBR322 to
CC give plasmid p3010. This is inserted into expression vector pKK223-3 to
CC give plasmid p4014. This is used to transform E.coli. It produces
CC desulphatohirudin, which is active as a thrombin inhibitor. It allows
CC easy mass prodn.
XX
XX Sequence 208 BP; 82 A; 23 C; 43 G; 60 T; 0 other;
SQ

```

Query Match 27.4%; Score 30.4; DB 10; Length 208;  
Best Local Similarity 77.1%; Pred. No. 0.4;  
Matches 37; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

Qy 1 tctagaggcgaaataatcaatcggttactgcggaagggtactccacaa 48
   ||| || ||||| ||||| ||||| || ||||| || ||| ||
Db 100 tctgatggagaaaaataatcggttactgagaggaggaacaccgaaa 147

```

Search completed: July 15, 2002, 22:44:48  
Job time: 10061 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:31:48 ; Search time 174.99 Seconds  
(without alignments)  
155.811 Million cell updates/sec

Title: US-10-053-641-3

Perfect score: 111

Sequence: 1 tctagaggcgaaataatca.....agaacctgcaataataggc 111

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PCUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	33.3	195	1	US-08-044-506B-27
2	35.8	32.3	563	1	US-08-044-506B-22
C	35.2	31.7	128	1	US-08-262-384A-5
4	35.2	31.7	235	1	US-08-262-384A-7
C	35.2	31.7	235	1	US-08-262-384A-8
6	35.2	31.7	726	1	US-08-262-384A-9
C	35.2	31.7	726	1	US-08-262-384A-10
8	34	30.6	106	1	US-08-262-384A-4
9	33	29.7	59	1	US-07-859-453E-5
10	33	29.7	59	1	US-07-859-453E-9
11	32	28.8	279	1	US-08-186-222-3
12	31.4	28.3	59	1	US-07-859-453E-11
13	31	27.9	195	1	US-08-406-948A-7
14	30.8	27.7	212	6	5180668-10
15	30.8	27.7	227	2	US-07-982-064-8
C	30.8	27.7	8491	2	US-08-757-439-1
17	30.4	27.4	67	6	5422249-9
18	30.4	27.4	122	6	5422249-8
19	30.4	27.4	195	6	5166318-14
20	30.4	27.4	217	6	5422249-13
21	30.4	27.4	276	6	5166318-15
22	30.4	27.4	526	6	5166318-16
23	29.8	26.8	43	1	US-07-985-110-9
24	29.8	26.8	43	1	US-08-099-053-9
25	29.8	26.8	43	1	US-08-452-829-9
26	29.4	26.5	1082	1	US-08-715-252-1
27	29.4	26.5	1082	2	US-08-453-051-3

28	29.4	26.5	1130	1	US-08-406-948A-5	Sequence 5, Appl
29	28.8	25.9	198	1	US-07-910-528-2	Sequence 2, Appl
30	28.8	25.9	198	1	US-08-348-972-2	Sequence 2, Appl
31	28.8	25.9	198	2	US-08-861-459-3	Sequence 3, Appl
32	28.6	25.8	59	1	US-07-859-453B-7	Sequence 7, Appl
33	28.2	25.4	201	1	US-07-854-596B-1	Sequence 1, Appl
34	28.2	25.4	223	1	US-07-854-596B-7	Sequence 7, Appl
35	28.2	25.4	420	1	US-07-854-596B-8	Sequence 8, Appl
36	28.2	25.4	1458	1	US-07-854-596B-42	Sequence 42, Appl
37	28.2	25.4	1467	1	US-07-854-596B-46	Sequence 46, Appl
38	26.6	24.0	46	2	US-08-861-459-7	Sequence 7, Appl
39	26.2	23.6	2115	2	US-08-767-026-3	Sequence 3, Appl
40	26	23.4	69	2	US-08-560-098A-35	Sequence 35, Appl
41	26	23.4	227	6	5164304-3	Patent No. 5164304
C	42	26	1983	1	US-08-221-817-21	Sequence 21, Appl
C	43	26	1983	1	US-08-454-439-21	Sequence 21, Appl
C	44	26	1983	5	PCT-US94-10487-21	Sequence 21, Appl
C	45	25.6	23.1	35	US-07-985-110-10	Sequence 10, Appl

#### ALIGNMENTS

RESULT 1

US-08-044-506B-27

; Sequence 27, Application US/08044506B

; Patent No. 5552299

; GENERAL INFORMATION:

; APPLICANT: OTT, Istvan

; APPLICANT: KLUPP, Tibor

; APPLICANT: MOLNAR, Istvan

; APPLICANT: PATTHY, Andras

; APPLICANT: BARTA, Istvan

; APPLICANT: BARKO nee TOOTH, Zsuzsa

; APPLICANT: AMBRUS, Gabor

; APPLICANT: SALAT, Janos

; APPLICANT: TEGDES, Aniko

; APPLICANT: MORAVCSIK, Imre

; TITLE OF INVENTION: Plasmids and Process for Producing

; TITLE OF INVENTION: Recombinant Desulphatohirudin HV-1 Peptides

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Keil & Weinkauf

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage

; COMPUTER: IBM AT-compatible, 80286 processor

; OPERATING SYSTEM: MS-DOS version 5.0

; SOFTWARE: Wordperfect version 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/044,506B

; FILING DATE: 09-APR-1993

; CLASSIFICATION: 435

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 195 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-044-506B-27

Query Match 33.3%; Score 37; DB 1; Length 195;

Best Local Similarity 85.1%; Pred. No. 0.00017;

Matches 40; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 tctagaggcgaaataatcaatcgcttactgaggaagtactccaca 47

|||: ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 94 TCTRACGGTGAAAAAATCAATGTGTCACTGGCGAAGGTACTCCAAA 140

RESULT 2  
US-08-044-506B-22  
; Sequence 22, Application US/08044506B  
; Patent No. 5552299  
; GENERAL INFORMATION:  
; APPLICANT: OTT, Istvan  
; APPLICANT: KLUPP, Tibor  
; APPLICANT: MOLNAR, Istvan  
; APPLICANT: PATTY, Andras  
; APPLICANT: BARTHA, Istvan  
; APPLICANT: BARKO nee TOTH, Zsuzsa  
; APPLICANT: AMBRUS, Gabor  
; APPLICANT: SALAT, Janos  
; APPLICANT: TEGDES, Aniko  
; APPLICANT: MORAVCSIK, Imre  
; TITLE OF INVENTION: Recombinant Desulphatohirudin HV-1 Peptides  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
; COMPUTER: IBM AT-compatible, 80286 processor  
; OPERATING SYSTEM: MS-DOS version 5.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/044, 506B  
; FILING DATE: 09-APR-1993  
; CLASSIFICATION: 435  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 563 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-044-506B-22

Query Match 32.3%; Score 35.8; DB 1; Length 563;  
Best Local Similarity 85.1%; Pred. No. 0.00065;  
Matches 40; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 1 tctagagcgagaaataatcggttactgaggaaggtactccaca 47  
Db 442 TCTGACGTGAAATAATCAATGTGTCTGCGGAGGTACTCCAAA 488

RESULT 3  
US-08-262-384A-5/C  
; Sequence 5, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0

; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262, 384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 128 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-262-384A-5

Query Match 31.7%; Score 35.2; DB 1; Length 128;  
Best Local Similarity 83.3%; Pred. No. 0.00065;  
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 tctagagcgagaaataatcggttactgaggaaggtactccaca 48  
Db 113 TCTGACGGCGAAAAAACCAGTGTCTGCGGAGGTACTCCCGAAA 66

RESULT 4  
US-08-262-384A-7  
; Sequence 7, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262, 384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 235 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-262-384A-7









RESULT 14  
5180668-10  
; Patent No. 5180668  
; APPLICANT: CRAUSE, PETER; HABERMANN, PAUL; TRIPIER, DOMINIQUE  
; TITLE OF INVENTION: HIRUDIN DERIVATIVE  
; NUMBER OF SEQUENCES: 10  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/295,422  
; FILING DATE: 10-JAN-1989  
; SEQ ID NO: 10  
; LENGTH: 212  
5180668-10

Query Match	27.7%	Score 30.8;	DB 2;	Length 227;
Best Local Similarity	83.3%	Pred. No. 0.03;		
Matches 35;	Conservative 0;	Mismatches 7;	Indels	
Qy	7	ggcgaataaatcaatcgcttactggcgaggttactccacaa	48	
Db	97	GGTGAAGAAGAACCAAGTCGGTTACTGGCGAAGTACCCTCGAAA	138	

Search completed: July 15, 2002, 22:31:48  
Job time: 18474 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:48:48 ; Search time 6165.88 Seconds  
(without alignments)  
242.976 Million cell updates/sec

Title: US-10-053-641-3  
Perfect score: 111  
Sequence: 1 tctaggaggcgaaaaataca.....agaacctgcataatagggc 111

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estcov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30.8	27.7	455	AQ704885	AQ704885 HS_5521_B
2	30.2	27.2	346	D48127	D48127 RICS14185A
3	30	27.0	849	CNS08YL0	AL421162 T7 end of
4	29.6	26.7	369	D47621	D47621 RICS13224A
5	29	26.1	167	BH005024	BH005024 BMBAC08P1
6	29	26.1	241	BG907200	BG907200 Tatr1157H
7	29	26.1	287	BI233876	BI233876 949037C04
8	29	26.1	409	AJ234149	AJ234149 B927 133
9	29	26.1	421	T76275	T76275 11053 Lambd
10	29	26.1	442	BI478547	BI478547 949070G02
11	29	26.1	551	BG316445	BG316445 947023E01
12	29	26.1	566	BI233858	BI233858 949037A03
13	28.8	25.9	357	TA62H05P	AL465105 T. brucei
14	28.8	25.9	553	TA278H10Q	AL486158 T. brucei
15	28.8	25.9	573	TA174F12Q	AL474540 T. brucei
16	28.8	25.9	660	AL554638	AL554638 AL554638
17	28.6	25.8	328	D47040	D47040 RICS12113A

C 18	28.6	25.8	343	10	D46970	D46970 RICS11963A
C 19	28.4	25.6	463	10	T21551	T21551 3559 Lambda
C 20	28.4	25.6	523	9	AA855282	AA855282 vw70G01.r
C 21	28.4	25.6	655	10	BI946448	BI946448 bES744.5'
C 22	28.2	25.4	287	10	D47557	D47557 RICS11130A
C 23	28.2	25.4	499	10	BE411902	BE411902 ISC010.E0
C 24	28.2	25.4	502	10	R90001	R90001 16356 Lambd
C 25	28.2	25.4	927	10	BI106318	BI106318 602890643
C 26	28	25.2	405	10	D48111	D48111 RICS14162A
C 27	28	25.2	453	10	T44933	T44933 8196 Lambda
C 28	28	25.2	480	9	AA873427	AA873427 OH77d11.s
C 29	28	25.2	507	12	BH504679	BH504679 BOHQ40TF
C 30	28	25.2	525	10	N38188	N38188 19415 Lambd
C 31	28	25.2	529	10	N65210	N65210 20250 Lambd
C 32	28	25.2	537	10	H37346	H37346 15475 Lambd
C 33	28	25.2	545	10	H77106	H77106 17537 Lambd
C 34	28	25.2	548	10	N37528	N37528 18755 Lambd
C 35	28	25.2	570	10	BF317079	BF317079 601903478
C 36	28	25.2	603	10	BE468512	BE468512 lPHOKF004
C 37	28	25.2	651	10	BM359653	BM359653 GA_Ea002
C 38	28	25.2	915	12	CNS034R1	AL227782 Tetraodon
C 39	27.8	25.0	334	9	AV527342	AV527342 AV527342
C 40	27.8	25.0	342	9	AV528031	AV528031 AV528031
C 41	27.8	25.0	356	9	BE230770	BE230770 99MJ388.R
C 42	27.8	25.0	377	9	AV526530	AV526530 AV526530
C 43	27.8	25.0	422	12	AQ631826	AQ631826 RPI-11-4
C 44	27.8	25.0	463	10	T04150	T04150 100 Lambda-
C 45	27.8	25.0	465	9	AV526935	AV526935 AV526935

#### ALIGNMENTS

#### RESULT 1

AQ704885 455 bp DNA linear GSS 07-JUL-1999  
HS\_5521\_B2\_B09\_T7A RPI-11 Human Male BAC Library Homo sapiens  
genomic clone plate-1097 Col-18 Row=D, DNA sequence.

AQ704885  
GI:5414399

SEQUENCE-Tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPI-11. For BAC library availability, please contact Pletier de Jong (pletier@u.washington.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Research Genetics (Inforesgen.com). BAC end Web Server: http://www.htsc.washington.edu

Plate: 1097 row: D column: 18  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 455.  
Location/Qualifiers  
1. .455  
/organism="Homo sapiens"

FEATURES  
source

```

/db_xref="taxon:9606"
/clone="Plate:1097 Col=18 Row=D"
/clone_lib="RPI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      202 a  101 c  59 g   92 t   1 others
ORIGIN

Query Match      27.7%; Score 30.8; DB 12; Length 455;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 44; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 39 tactccacaggtttgaatcccggaagaatagaacccgagctctccgacgacgaacct 98
Db 243 TTCTCAACAATTTTAAACCGAAATTTATACCAACCACTTCTCAGACCACACACAA 302
QY 99 gcaata 104
Db 303 TAAATA 308

RESULT 2
D48127/c
LOCUS      D48127      346 bp  mRNA  linear  EST 02-AUG-1995
DEFINITION R1CS14185A Rice green shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION  D48127
VERSION     D48127.1 GI:701836
KEYWORDS   EST.
SOURCE     Oryza sativa.
ORGANISM   Oryza sativa
REFERENCE  1 (bases 1 to 346)
AUTHORS    Sasaki, T., Miyao, A. and Yamamoto, K.
TITLE      Rice cDNA from callus 1995
JOURNAL    Unpublished (1995)
COMMENT    National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/.
FEATURES
source
1..346
Location/Qualifiers
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone_lib="Rice green shoot"
/note="Green shoot (8 days old)"
BASE COUNT      61 a   95 c  140 g   44 t   6 others
ORIGIN

Query Match      27.2%; Score 30.2; DB 10; Length 346;
Best Local Similarity 63.8%; Pred. No. 18;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 34 gaaggtactccacaggtttgaatcccggaagaatagaacccgagctctccgacgacg 93
Db 303 GAAGGTACTCCGCGGCTTCCCNAGCGCGAAGCGGTGCGAGCGGTAGTCGCGCAGCGG 244
QY 94 aacctgcaa 102
Db 243 ANCGGTCCA 235

```

```

RESULT 3
CNS06YL0
LOCUS      CNS06YL0      849 bp  DNA  linear  GSS 06-JUL-2001
DEFINITION T7 end of clone AY0AA010E04 of library AY0AA from strain CBS 6340
of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION  AL421162
VERSION     AL421162.1 GI:12204361
KEYWORDS   GSS.
SOURCE     Kluyveromyces thermotolerans.
ORGANISM   Kluyveromyces thermotolerans
REFERENCE  1 (bases 1 to 849)
AUTHORS    Malpertuy, A., Llorente, B., Blandin, G., Artiguenave, F., Wincker, P.
and Dujon, B.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 10.
JOURNAL    FEBS Lett. 487 (1), 61-65 (2000)
MEDLINE    20584720
REFERENCE  2 (bases 1 to 849)
AUTHORS    Souciet, J.-L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neugeglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL    FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE    20584711
REFERENCE  3 (bases 1 to 849)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT    This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Pichia
lactis var. lactis, Kluyveromyces thermotolerans, Kluyveromyces
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
1..849
Location/Qualifiers
/organism="Kluyveromyces thermotolerans"
/strain="CBS 6340"
/db_xref="taxon:4916"
/clone="AY0AA010E04"
/clone_lib="AY0AA"
/note="end : T7"
BASE COUNT      177 a  245 c  230 g  190 t   7 others
ORIGIN

Query Match      27.0%; Score 30; DB 12; Length 849;
Best Local Similarity 55.9%; Pred. No. 24;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 5 gagggcaaaaaaatcaatcgcttactgaggaaggtactccacgaaggttgaatcccgga 64
Db 13 GTGGTGTGAGAAGGGATAGAGCTCTCTTGAAGGGCCCTCCGCAATACAGGGAAGAGTA 72
QY 65 aqaatagaacgcagctctccgacgacgaacacctgcaataat 106
Db 73 CGGTGAGAAATAGCAGCTCTCCGCTCTTAAGTACTGACTGATTGAAT 114

RESULT 4
D47621/c

```



LOCUS D47621 369 bp mRNA linear EST 02-AUG-1995  
 DEFINITION RGS1222A Rice green shoot Oryza sativa cDNA, mRNA sequence.  
 ACCESSION D47621  
 VERSION D47621.1 GI:701330  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.  
 1 (bases 1 to 369)  
 TITLE Sasaki, T., Miyao, A. and Yamamoto, K.  
 JOURNAL Rice cDNA from callus 1995  
 COMMENT Unpublished (1995)  
 Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@br.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/.  
 FEATURES  
 Location/Qualifiers  
 1..369  
 /organism="Oryza sativa"  
 /strain="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone\_lib="Rice green shoot"  
 /note="Green shoot (8 days old)"  
 62 a 107 c 149 g 47 t 4 others  
 BASE COUNT 62 a 107 c 149 g 47 t 4 others  
 ORIGIN  
 Query Match 26.7%; Score: 29.6; DB 10; Length 369;  
 Best Local Similarity 63.8%; Pred. No. 28;  
 Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
 QY 34 gaagctactcccaaggcttgatccgcggaagataagaacccagctctccgacgacg 93  
 Db 284 GNAGGTACTCCCGCGCTTCCCGAGGCCGACGCGGTGCGAGCGGTGCGCGAGCGG 225  
 QY 94 aacctgcaa 102  
 Db 224 AACCGTCCA 216  
 RESULT 5  
 BH005024  
 LOCUS BMBAC09P14SP6.PSU Brugia malayi Genomic BAC Library 1 & 2 Brugia  
 DEFINITION malayi genomic, DNA sequence.  
 ACCESSION BH005024  
 VERSION BH005024.1 GI:13946100  
 KEYWORDS GSS.  
 SOURCE Brugia malayi.  
 ORGANISM Brugia malayi  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
 Onchocercidae; Brugia.  
 1 (bases 1 to 167)  
 TITLE Whitton, C., Daub, J., Ware, J., Quail, M., Hall, N., Barrell, B., Foster  
 J., Guilianno, D., Slatko, B. and Blaxter, M.  
 JOURNAL Genome survey sequences from the human parasitic nematode Brugia  
 malayi  
 COMMENT Unpublished (2000)  
 Contact: Blaxter ML  
 Institute of Cell, Animal and Population Biology  
 University of Edinburgh  
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
 3JT, UK  
 Tel: +44 131 650 6760  
 Fax: +44 131 670 5450  
 Email: mark.blaxter@ed.ac.uk  
 Library constructed by Jesse Pope-Chappel and Jeremy Foster. The

sequence was generated by The Pathogen Sequencing Unit, The Sanger  
 Centre, Cambridge, UK in collaboration with Mark Blaxter, ICAPP  
 University of Edinburgh, Edinburgh, UK  
 Seq primer: SP6 (CGCCAGCTATTAGGTGACAC)  
 Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
 1..167  
 /organism="Brugia malayi"  
 /strain="TRS"  
 /db\_xref="taxon:6279"  
 /clone\_lib="Brugia malayi Genomic BAC Library 1 & 2"  
 /sex="Mixed (male and female)"  
 /issue\_type="whole parasite"  
 /dev\_stage="adult"  
 /note="Vector: pBelobAC II; Site 1: Hind III; Brugia  
 malayi genomic DNA was partially cleaved with Hind III and  
 size fractionated. 18,000 clones were generated from 2  
 libraries with mean insert size 60 kbp. The library was  
 constructed by Jesse Pope-Chappel, Smith College  
 Northampton MA and Dr Jeremy Foster, New England Biolabs,  
 MA."  
 76 a 22 c 23 g 46 t  
 BASE COUNT 76 a 22 c 23 g 46 t  
 ORIGIN

Query Match 26.1%; Score 29; DB 12; Length 167;  
 Best Local Similarity 61.0%; Pred. No. 38;  
 Matches 47; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 3 tagagcgcaaaaataatcgcttactgcggaagtgactcccaaggtcttgaaatccg 62  
 Db 6 TAGAAGCTATATAAATAAAGAAACATTGTAGATTGTCTCCAAAGAAATTCAAACGAG 65

QY 63 gaagaatagaacccgca 79  
 Db 66 CAACAGAGAGAAAGTACA 82

RESULT 6  
BG07200/c

LOCUS TaLR1157H06R TaLR1 Triticum aestivum cDNA clone TaLR1157H06 5',  
 DEFINITION mRNA sequence.

ACCESSION BG07200  
 VERSION BG07200.1 GI:14314876  
 KEYWORDS EST.  
 SOURCE Bread wheat.  
 ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
 ; Triticeae; Triticum.  
 1 (bases 1 to 241)  
 TITLE Cloutier, S., Dong, G. and Walsh, A.  
 JOURNAL Wheat functional genomics- Thatcher Lr1 cDNA library  
 COMMENT Unpublished (2001)  
 Contact: Dr. Sylvie Cloutier  
 Cereal Research Centre, Agriculture and Agri-food Canada  
 195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9  
 Tel: (204) 983-2340  
 Fax: (204) 983-4604  
 Email: scloutier@em.agr.ca

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

was cloned directionally, not all sequences generated with reverse  
 primer were from the 5' end (same with forward primer and 3' end).  
 Average insert size is >2.2 kb  
 Plate: 157 row: H column: 06  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..241  
 /organism="Triticum aestivum"  
 /cultivar="Thatcher Lr1"  
 /db\_xref="taxon:4565"  
 /clone\_lib="TaLR1"  
 /clone\_lib="TaLR1"

FEATURES  
source

Location/Qualifiers  
 1..241  
 /organism="Triticum aestivum"  
 /cultivar="Thatcher Lr1"  
 /db\_xref="taxon:4565"  
 /clone\_lib="TaLR1"  
 /clone\_lib="TaLR1"

```

/tissue_type="Leaf tissue"
/dev_stage="14 Days old"
/lab_host="E. coli XL0LR"
/notes="Vector: Lambda ZapII; mass excised in plasmid
vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI;
mRNA obtained from wheat NIL Thatcher Lrl 24 hours after
inoculation with leaf rust pathogen Puccinia triticina
race BBB carrying the avirulence gene Avr1."
BASE COUNT      37 a      89 c      84 g      31 t
ORIGIN

Query Match      26.1%; Score 29; DB 10; Length 241;
Best Local Similarity 63.8%; Pred. No. 40;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 34 gaaggctcccaagcgtttgaatcccggaagaatagaacccgagctctccgacgacag 93
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 231 GCAGGTACTCCGGGGCTTCCCAAGGCCGAGGGGTAGAGCCGTAGTCGCGCAGCAGCG 172
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 94 aacctgcaa 102
| | | | |
Db 171 TCCCGTCGA 163
| | | | |

RESULT 7
LOCUS B1233876/c
DEFINITION 949037C04.y2 949 - Juvenile leaf and shoot cDNA from Steve Moose
zeae may cDNA, mRNA sequence.
ACCESSION B1233876
VERSION B1233876
KEYWORDS B1233876.1 GI:14701458
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE 1 (bases 1 to 287)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL University
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949037 row: C column: 04.
FEATURES
Source
Location/Qualifiers
1..287
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/tissue_type="immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/notes="Organ: juvenile vegetative shoots; Vector:
pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridap vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing; includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf

```

```

primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."
BASE COUNT      39 a      99 c      105 g      44 t
ORIGIN

Query Match      26.1%; Score 29; DB 10; Length 287;
Best Local Similarity 63.8%; Pred. No. 42;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 34 gaaggctcccaagcgtttgaatcccggaagaatagaacccgagctctccgacgacag 93
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 274 GCAGGTACTCCAGCGGCTTCCCGAGCCGACGCGGTGCAACCCCTAGTCGCGCAGCAGCG 215
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 94 aacctgcaa 102
| | | | |
Db 214 ACCCGTCGA 206
| | | | |

RESULT 8
LOCUS AJ234149
DEFINITION B927 133 Trypanosoma brucei strain 927 ESTs Trypanosoma brucei cDNA
, mRNA sequence.
ACCESSION AJ234149
VERSION AJ234149.1 GI:3790347
KEYWORDS EST.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
REFERENCE 1 (bases 1 to 409)
AUTHORS van Deursen,F.J., Shahi,S.K., Turner,C.M.R., Hartmann,C.,
Guerra-Giraldez,C., Matthews,K.R. and Clayton,C.E.
TITLE Characterisation of the growth and differentiation in vivo and in
vitro of bloodstream-form Trypanosoma brucei strain TREU 927
JOURNAL Mol. Biochem. Parasitol. 112 (2), 163-171 (2001)
MEDLINE 21126580
COMMENT Contact: Shahi SK
ZMBH (Centre for Molecular Biology)
Heidelberg University
Im Neuenheimer Feld 282, 69120 Heidelberg, Germany
POLYA-No.
FEATURES
Source
Location/Qualifiers
1..409
/organism="Trypanosoma brucei"
/strain="927"
/db_xref="taxon:5691"
/clone_lib="Trypanosoma brucei strain 927 ESTs"
BASE COUNT      83 a      95 c      120 g      110 t      1 others
ORIGIN

Query Match      26.1%; Score 29; DB 9; Length 409;
Best Local Similarity 58.8%; Pred. No. 44;
Matches 50; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 23 gcgttactgcgaaggtactcccaaggtttgaatcccggaagaatagaacccgagctc 82
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GTGGAACGAGCAGCATTTACTCCACAGGAGTTGATTAAAGGAAGAAATCTCCGACGAGCT 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 83 tcgcgacgacagaacctgcaataata 107
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TGCATGCACATACGGCTCCTTGATA 85
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
LOCUS T76275/c
DEFINITION T76275
T76275 421 bp mRNA linear EST 09-JAN-1998
11053 Lambda-PRL2 Arabidopsis thaliana cDNA clone 148P10T7, mRNA
sequence.
ACCESSION T76275
VERSION T76275.1 GI:935283

```

KEYWORDS  
SOURCE

EST.  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE  
1 (bases 1 to 421)  
AUTHORS  
Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh,  
L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel,  
E., and Somerville, C.  
TITLE  
Genes galore: a summary of methods for accessing results from  
large-scale partial sequencing of anonymous Arabidopsis cDNA clones  
Plant Physiol. 106, 1241-1255 (1994)  
95148729  
On Apr 14, 1993 this sequence version replaced gi:693037.  
Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI  
Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 22313tcn@bm.cl.msu.edu  
Seq primer: T7 dye primer.  
FEATURES  
Location/Qualifiers  
1..421  
/organism="Arabidopsis thaliana"  
/strain="var columbia"  
/db\_xref="taxon:3702"  
/clone="148P107"  
/clone\_lib="Lambda-PRL2"  
/note="Vector: lambda Zip-Lox; Site.1: Sal; Site.2: Not;  
Lambda PRL2 is a cDNA library derived from equal  
quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
day germinated etiolated seedlings; 2) tissue culture  
grown roots; 3) staged plants half with 24 hour light  
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
same plants as 3 but aerial tissue (stems, flowers and  
siliques. The vector is BRL's lambda Zip-Lox. The cDNA  
inserts were directionally cloned with Sal-Not arms using  
oligo dt primed cDNA."

JOURNAL  
MEDLINE  
COMMENT

On Apr 14, 1993 this sequence version replaced gi:693037.  
Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI  
Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 22313tcn@bm.cl.msu.edu  
Seq primer: T7 dye primer.  
FEATURES  
Location/Qualifiers  
1..421  
/organism="Arabidopsis thaliana"  
/strain="var columbia"  
/db\_xref="taxon:3702"  
/clone="148P107"  
/clone\_lib="Lambda-PRL2"  
/note="Vector: lambda Zip-Lox; Site.1: Sal; Site.2: Not;  
Lambda PRL2 is a cDNA library derived from equal  
quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
day germinated etiolated seedlings; 2) tissue culture  
grown roots; 3) staged plants half with 24 hour light  
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
same plants as 3 but aerial tissue (stems, flowers and  
siliques. The vector is BRL's lambda Zip-Lox. The cDNA  
inserts were directionally cloned with Sal-Not arms using  
oligo dt primed cDNA."

BASE COUNT  
ORIGIN

106 a 101 c 102 g 96 t 16 others

Query Match  
Best Local Similarity  
Matches

26.1%; Score 29; DB 10; Length 421;  
62.0%; Pred. No. 44;  
44; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 34 gaaggtactccacaagcttgaatcccggaagaataagaacccgagtcctccgacgacag 93

|||||  
Db 284 GGAGATACCTCCCGGTTTGCTTAACCAAGAGTCAACCCGTAGTCCCGACAATG 225

QY 94 aacctgcaata 104

|||||  
Db 224 AACCGTCGANA 214

RESULT 10  
BI478547/c  
LOCUS

BI478547 442 bp mRNA linear EST 28-AUG-2001  
949070G02.y1 949 - Juvenile leaf and shoot cDNA from Steve Moose  
zea mays cDNA, mRNA sequence.

DEFINITION  
ACCESSION  
BI478547

BI478547.1 GI:15323491  
KEYWORDS  
SOURCE  
zea mays.  
zea mays.  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
Clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 442)  
AUTHORS  
Walbot, V.

REFERENCE  
AUTHORS

Walbot, V.

TITLE  
JOURNAL  
COMMENT

Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 949070 row: G column: 02.  
Location/Qualifiers  
1..442  
/organism="Zea mays"  
/cultivar="W64A"  
/db\_xref="taxon:4577"  
/clone\_lib="949 - Juvenile leaf and shoot cDNA from Steve  
Moose"  
/tissue\_type="immature leaf primordium and vegetative  
meristem"  
/dev\_stage="4 stages from 3-13 days after imbibing"  
/lab\_host="E. coli XL0LR"  
/note="Organ: juvenile vegetative shoots; Vector:  
pAD-GAL4-2.1; Site.1: EcoRI; Site.2: XhoI; Equal amounts  
of total RNA by weight from 4 tissue sources (see below)  
were pooled, polyA+ RNA isolated, and cDNA synthesized for  
EcoRI (5') and XhoI (3') directional cloning into lambda  
Hybridzap vector from Stratagene. Tissue Sources: 1. Whole  
shoots 3 days after sowing/imbibing in wet soil. 2. Basal  
1.5 cm shoots 6 days after sowing - includes yellow  
portions of developing leaves 1-5, primordia from 6-8, and  
the vegetative apex. 3. Non-green portions of developing  
leaves 4-5 and the vegetative apex. Including adult leaf  
primordia, 9 days after sowing. 4. Partially expanded and  
greening leaves 4-5 at 13 days after sowing."

FEATURES  
source

BASE COUNT 67 a 164 c 144 g 67 t

Query Match  
Best Local Similarity  
Matches

26.1%; Score 29; DB 10; Length 442;  
63.8%; Pred. No. 45;  
44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 34 gaaggtactccacaagcttgaatcccggaagaataagaacccgagtcctccgacgacag 93

|||||  
Db 245 GCAGGTACTCCAGGGGCTTCGCCGACGCCGACGGTGCAGACCGTAGTCCCGACCGCG 186

QY 94 aacctgcaa 102

|||||  
Db 185 AGCGTCGA 177

RESULT 11  
BG316445/c  
LOCUS

BG316445 551 bp mRNA linear EST 26-FEB-2001  
947023E01.y1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA  
sequence.

ACCESSION  
BG316445

BG316445.1 GI:13125875  
KEYWORDS  
SOURCE  
zea mays.  
zea mays.  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
Clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 551)  
AUTHORS  
Walbot, V.

REFERENCE  
AUTHORS

Walbot, V.

TITLE  
JOURNAL  
COMMENT

Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University



```

Query Match      25.9%; Score 28.8; DB 12; Length 357;
Best Local Similarity 54.8%; Pred. No. 50;
Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 tctagagcgcaaaaaatcaatcgcttactgcggaaggctactccacaaggcttgaatcc 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 41 TATGGGGAAGGACTAATAATGAATAGTTAGGACGAGGGAATATTAGTTGCATTAACTCG 100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 cggagaagaatagaacccgagctctccgacgacagaacctgcgaata 104
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 101 CCACAGAAAAAAGTGGCCCGGACCAACATCTCTTCACTA 144
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
LOCUS TA278H10Q 553 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 278h10, reverse sequence,
            genomic survey sequence.
ACCESSION AL486158
VERSION AL486158.1 GI:11851971
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
          Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
          Trypanosoma.
REFERENCE 1 (bases 1 to 553)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
        Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
        Melville,S.E., Rajandream,M.A. and Barrell,B.G.
        Direct Submission
        Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
        project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
        Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
        nh@sanger.ac.uk
        Constructed at the Institute for Genomic Research (TIGR),
        Rockville, MD. Genomic DNA isolated from a cloned population of
        Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
        to give a tight size distribution (
        4 kb). The v + i method used for the library construction is
        described in detail in Smith, H. and Venter, J.C. (Making small
        insert libraries for whole genome shotgun sequencing projects. In
        Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
        Barrell, Oxford University Press, 1999).
        Email: nelsayed@tigr.org
        Details of T. brucei sequencing at the Sanger Centre are available
        at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
    source
        1..553
        /organism="Trypanosoma brucei"
        /strain="TREU927"
        /db_xref="taxon:5691"
        /clone="278h10"
BASE COUNT 199 a 92 c 99 g 163 t
ORIGIN

Query Match      25.9%; Score 28.8; DB 12; Length 553;
Best Local Similarity 54.8%; Pred. No. 54;
Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 tctagagcgcaaaaaatcaatcgcttactgcggaaggctactccacaaggcttgaatcc 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 398 TATGGGGAAGGACTAATAATGAATAGTTAGGACGAGGGAATATTAGTTGCATTAACTCG 457
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 cggagaagaatagaacccgagctctccgacgacagaacctgcgaata 104
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 458 CCACAGAAAAAAGTGGCCCGGACCAACATCTCTTCACTA 501
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
LOCUS TA174F12Q/c 573 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 174f12, reverse sequence,
            genomic survey sequence.
ACCESSION AL474540
VERSION AL474540.1 GI:11839754
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
          Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
          Trypanosoma.
REFERENCE 1 (bases 1 to 573)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
        Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
        Melville,S.E., Rajandream,M.A. and Barrell,B.G.
        Direct Submission
        Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
        project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
        Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
        nh@sanger.ac.uk
        Constructed at the Institute for Genomic Research (TIGR),
        Rockville, MD. Genomic DNA isolated from a cloned population of
        Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
        to give a tight size distribution (
        4 kb). The v + i method used for the library construction is
        described in detail in Smith, H. and Venter, J.C. (Making small
        insert libraries for whole genome shotgun sequencing projects. In
        Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
        Barrell, Oxford University Press, 1999).
        Email: nelsayed@tigr.org
        Details of T. brucei sequencing at the Sanger Centre are available
        at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
    source
        1..573
        /organism="Trypanosoma brucei"
        /strain="TREU927"
        /db_xref="taxon:5691"
        /clone="174f12"
BASE COUNT 134 a 123 c 112 g 204 t
ORIGIN

Query Match      25.9%; Score 28.8; DB 12; Length 573;
Best Local Similarity 54.8%; Pred. No. 54;
Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 tctagagcgcaaaaaatcaatcgcttactgcggaaggctactccacaaggcttgaatcc 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 381 TATGGGGAAGGACTAATAATGAATAGTTAGGACGAGGGAATATTAGTTGCATTAACTCG 322
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 cggagaagaatagaacccgagctctccgacgacagaacctgcgaata 104
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 321 CCACAGAAAAAAGTGGCCCGGACCAACATCTCTTCACTA 278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: July 15, 2002, 21:48:54
Job time: 18885 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:28:44 ; Search time 2368.24 Seconds

(without alignments)  
954.323 Million cell updates/sec

Title: US-10-053-641-4

Perfect score: 108

Sequence: 1 ggcgcgcctatttcagg.....aacgcattgatttttttcgc 108

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query  
No. Score Match Length DB ID Description

1	95.4	88.3	128	6	I41185	I41185 Sequence 5
2	95.4	88.3	235	6	I41188	I41188 Sequence 8
3	94	87.0	106	6	I41184	I41184 Sequence 4
C 4	94	87.0	226	12	SYNHR	M26762 Synthetic h
C 5	94	87.0	235	6	I41187	I41187 Sequence 7
C 6	93	86.1	726	6	I41189	I41189 Sequence 9
C 7	93	86.1	726	6	I41190	I41190 Sequence 10
C 8	86.2	79.8	220	12	SYNHMED	M14964 Synthetic h
C 9	86.2	79.8	224	6	A00630	A00630 Artificial
C 10	86.2	79.8	224	6	A00631	A00631 Artificial
C 11	86.2	79.8	224	6	A04277	A04277 Artificial
C 12	86.2	79.8	224	6	A04278	A04278 Artificial
C 13	86.2	79.8	224	6	A07345	A07345 Synthetic D
C 14	86.2	79.8	224	6	A07346	A07346 Synthetic D
C 15	84.2	78.0	279	6	A01139	A01139 Fusion DNA
C 16	84.2	78.0	279	6	I26634	I26634 Sequence 3
C 17	83.2	77.0	212	6	A13382	A13382 DNA sequenc
C 18	83.2	77.0	212	6	A13383	A13383 DNA sequenc
C 19	83.2	77.0	212	6	A34619	A34619 Synthetic h
C 20	83.2	77.0	212	6	A34620	A34620 Synthetic h
C 21	83.2	77.0	212	6	E00711	E00711 Synthetic D
C 22	83.2	77.0	8491	6	AR031529	AR031529 Sequence
C 23	82.6	76.5	122	6	A34236	A34236 Synthetic h
C 24	82.6	76.5	122	6	A34237	A34237 Synthetic h
C 25	82.6	76.5	217	6	A03695	A03695 Nucleotide
C 26	82.6	76.5	217	6	A03696	A03696 Nucleotide
C 27	82.6	76.5	217	6	A34238	A34238 Synthetic d
C 28	82.6	76.5	217	6	A34239	A34239 Synthetic d
C 29	82.6	76.5	217	6	E00637	E00637 DNA encodin
C 30	81.4	75.4	4477	12	SYNEXVECHD	M88535 Expression
C 31	81.2	75.2	195	6	A03693	A03693 Nucleotide
C 32	81.2	75.2	195	6	A03694	A03694 Nucleotide
C 33	81.2	75.2	226	6	A04615	A04615 Nucleotide
C 34	77.6	71.9	211	6	E03003	E03003 DNA encodin
C 35	76	70.4	238	6	A04616	A04616 HVI gene. 8
C 36	75.8	70.2	223	6	A18064	A18064 hirudin typ
C 37	75.8	70.2	223	6	A19994	A19994 SEQ ID NO:
C 38	75.8	70.2	223	6	A19999	A19999 SEQ ID NO:
C 39	75.8	70.2	223	6	A20035	A20035 SEQ ID NO:
C 40	75.8	70.2	223	6	I13188	I13188 Sequence 7
C 41	75.8	70.2	420	6	A20000	A20000 SEQ ID NO:
C 42	75.8	70.2	420	6	I13189	I13189 Sequence 8
C 43	75.8	70.2	1467	6	A20030	A20030 SEQ ID NO:
C 44	75.8	70.2	1467	6	I13218	I13218 Sequence 46
C 45	75.4	69.8	201	6	A34242	A34242 Hirudin cod

ALIGNMENTS

RESULT 1	I41185	I41185	Sequence 5 from patent US 5624822.	128 bp	DNA	linear	PAT 13-MAY-1997
LOCUS	I41185	I41185	Sequence 5 from patent US 5624822.				
DEFINITION	I41185	I41185	Sequence 5 from patent US 5624822.				
ACCESSION	I41185	I41185	Sequence 5 from patent US 5624822.				
VERSION	I41185.1	GI:2081775					
KEYWORDS	Unknown.						
SOURCE	Unknown.						
ORGANISM	Unclassified.						
REFERENCE	1 (bases 1 to 128)						
AUTHORS	Koerwer, W.						
TITLE	Hirudin fusion proteins and preparation of hirudin						
JOURNAL	Patent: US 5624822-A 5 29-APR-1997;						
FEATURES	Location/Qualifiers						
source	1..128						
BASE COUNT	20 a 33 c 34 g 41 t						
ORIGIN	/organism="unknown"						

Query Match 88.3%; Score 95.4; DB 6; Length 128;  
Best Local Similarity 94.3%; Pred. No. 1.9e-22;













GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:48 ; Search time 758.37 seconds  
(without alignments)  
244.507 Million cell updates/sec

Title: US-10-053-641-4

Perfect score: 108

Sequence: 1 ggcgcctattattgcagg.....aacgcattgatttttttcgc 108

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	ID	Description
C 1	93	86.1	726 12	AAQ12379
C 2	85.4	79.1	224 7	AAQ12379
C 3	84.2	78.0	279 12	AAQ13831
C 4	83.8	77.6	468 16	AAQ3244
C 5	83.2	77.0	227 22	AAF61507
C 6	82.6	76.5	217 7	AAQ60355
C 7	82.6	76.5	217 8	AAQ70323
C 8	82.4	76.3	185 12	AAQ14926
C 9	77.6	71.9	210 12	AAQ10452

Sequence encoding  
Factor Xa-cleavabl  
Factor Xa-cleavabl  
Desulphatohirudin  
Desulphatohirudin  
Synthetic hirudin  
Hirudin HV-1. Syn  
Desulphatohirudin  
Synthetic DNA enco  
Synthetic HVI gene  
Synthetic hirudin  
CUPI promoter, PHO  
Yeast CUP1 promote  
pJDB207/GAPFL-YHR  
Leech hirudin mute  
Sequence encoding  
Sequence encoding  
Factor Xa-cleavabl  
Encodes hirudin de  
Sequence of DNA in  
Encodes hirudin de  
Secretion plasmid  
Olesein-hirudin fu  
Desulphatohirudin  
Sequence encoding  
Hirudin variant.  
Partial Hirudin HV  
DNA encoding the f  
RHV2-Phe3,Gl33,Ty  
DNA encoding hirud  
Sequence encoding  
Hirudin gene HV1.  
Sequence encoding  
Recombinant hirudi  
DNA encoding leech  
DNA encoding leech

#### ALIGNMENTS

RESULT 1  
AAQ12379/c  
ID AAQ12379 standard; DNA; 726 BP.  
XX AAQ12379;  
AC AAQ12379;  
XX  
DT 17-SEP-1991 (first entry)  
XX  
DE Hirudin peptide/Protein A fusion gene.  
XX  
KW anticoagulant; fusion protein; ds.  
XX  
FH Key Location/Qualifiers  
FT mat\_peptide 1..489  
FT /tag= a  
FT /product= Protein A  
FT mat\_peptide 529..720  
FT /tag= b  
FT /product= Hirudin  
FT misc\_RNA 490..528  
FT /tag= c  
FT /product= linker oligopeptide Y  
XX  
PN DE3942580-A.  
XX  
PD 27-JUN-1991.  
XX  
PE 22-DEC-1989; 89DE-3942580.  
XX  
PR 22-DEC-1989; 89DE-3942580.  
XX  
PA (BADI ) BASF AG.





CC with the Leu-hirudin (LH) ((Leu1-Thr2)-63-desulfato-hirudin) sequence  
 CC linked to the C-terminus of the signal sequence. (I) is an intermediate  
 CC in recombinant production of LH, a known antithrombotic. The specified  
 CC signal sequence may also be used for secretory expression of other  
 CC proteins. (II) is processed directly to LH and this, in native form,  
 CC secreted from E. coli in high yield. This results, both during  
 CC fermentation and subsequent purification, in a higher concentration of  
 CC hirudin, reducing costs of production. The specified signal sequences  
 CC provide more efficient secretion than known sequences. This sequence  
 CC encodes a fragment of the S. marcescens hirudin protein.  
 XX  
 SQ Sequence 227 BP; 61 A; 60 C; 58 G; 48 T; 0 other;

Query Match 77.0%; Score 83.2; DB 22; Length 227;  
 Best Local Similarity 87.5%; Pred. No. 4e-19;  
 Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 5 gccctattatgcaggtattcttcgggattcttcaagtcgcgcgtgttgagactg 64  
 DB 204 GCTCTATTACTGAAGGTATTCTCTCAGGATCTCTTCGAAGTCGCGGTATGAGACTG 145  
 QY 65 cggttcggagtaccttcgcagtaacgcattgatttttttcgc 108  
 DB 144 CGGTTTCGGGTACCTTCGCAGTAACGCACCTGTTCTTTTCAC 101

RESULT 6  
 AAN60355/c  
 ID AAN60355 standard; DNA; 217 BP.  
 XX  
 AC AAN60355;  
 XX  
 DT 20-JUN-1991 (first entry)  
 XX  
 DE Desulphatohirudin (II).  
 XX  
 KW Desulphatohirudin; antibodies; thrombin; ds.  
 XX  
 OS Synthetic.

Key Location/Qualifiers  
 CDS 10..207  
 FT /\*tag= a  
 FT /product= hirudin  
 XX  
 PN EP168342-A.  
 XX  
 PD 15-JAN-1986.  
 XX  
 PF 10-JUN-1985; 85EP-0810268.  
 XX  
 PR 14-JUN-1984; 84CH-0288284.  
 XX  
 PA (CIBA ) CIBA GEIGY AG.  
 XX  
 PL Liersch M, Rink H, Marki W, Grutter MG, Meyhack B;  
 XX  
 DR WPI; 1986-015589/03.  
 DR P-PSDB; AAP60395.  
 XX

DNA sequences coding for hirudin and derivs. - and expression  
 PT vectors, transformed cells, monoclonal antibodies and hybridomas,  
 PT useful as thrombin inhibitor.

PS Disclosure; Page 10; 123pp; German.

XX The sequence comprises an EcoRI restriction enzyme site at the  
 CC 5'-end and a BamHI restriction enzyme site at the 3'-end.  
 CC The sequence may be introduced into a vector for the transformation  
 CC of hosts, e.g. E. coli. The hirudin can thus be prepared on a  
 CC large scale. The product and its derivs. are thrombin inhibitors,  
 CC useful in anticoagulant therapy, esp. when injected at doses

CC of 0.01-0.05 mg/kg.

XX  
 SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match 76.5%; Score 82.6; DB 7; Length 217;  
 Best Local Similarity 90.7%; Pred. No. 6.3e-19;  
 Matches 88; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 12 tattgcaggtattcttcgggattcttcaagtcgcgcgtgttgagactcggttc 71  
 DB 209 TATGCAGGTATTCTTCGGGATTTCTTCGAAGTCACCGTCGTTGTGAGACTCGGTTTC 150  
 QY 72 gtagtaccttcgcagtaacgcattgatttttttcgc 108  
 DB 149 GGGTACCTTCACCGGTAAAGCACTGGTTTTTTTCAC 113

RESULT 7  
 AAN70323/c  
 ID AAN70323 standard; DNA; 217 BP.  
 XX  
 AC AAN70323;  
 XX  
 DT 02-APR-1991 (first entry)  
 XX  
 DE Sequence of the desulphatohirudin gene.  
 XX  
 KW Anticoagulant; thrombin inhibitor; ds.  
 XX  
 PN EP225633-A.  
 XX  
 PD 16-JUN-1987.  
 XX  
 PF 09-DEC-1986; 86EP-0117098.  
 XX  
 PR 29-MAY-1986; 86GB-0013088.  
 PR 12-DEC-1985; 85GB-0030631.  
 XX  
 PA (CIBA ) CIBA GEIGY AG.  
 PA (PLAN-) PLANTORGAN WERK HEINRICH.  
 PA (CHRI-) PLANTORGANW CHRISTENSEN.  
 XX  
 PI Meyhack B, Marki W, Heim J;  
 XX  
 DR WPI; 1987-164868/24.

New DNA constructs and hybrid vectors for transformation of yeast  
 etc. - useful for prodn. and secretion of protein with hirudin  
 activity for use as thrombin inhibitors.  
 XX  
 PS Example: p44; 146pp; English.  
 XX  
 CC The preferred DNA construct of the invention contains the PHO5  
 CC promoter and a DNA segment consisting of the PHO5 signal sequence  
 CC upstream of and in reading frame with a DNA sequence coding for  
 CC mature desulphatohirudin. The segment is under the transcriptional  
 CC control of the PHO5 promoter and the 3' flanking sequence of the  
 CC PHO5 gene.  
 XX  
 SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match 76.5%; Score 82.6; DB 8; Length 217;  
 Best Local Similarity 90.7%; Pred. No. 6.3e-19;  
 Matches 88; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 12 tattgcaggtattcttcgggattcttcaagtcgcgcgtgttgagactcggttc 71  
 DB 209 TATGCAGGTATTCTTCGGGATTTCTTCGAAGTCACCGTCGTTGTGAGACTCGGTTTC 150  
 QY 72 gtagtaccttcgcagtaacgcattgatttttttcgc 108



Db 149 GGGGTACCTTCACCGGTAAACGACTGGTTTTTTCAC 113

## RESULT 8

AAQ14926/c  
ID AAQ14926 standard; DNA; 185 BP.

XX AC AAQ14926;  
XX  
XX DT 24-FEB-1992 (first entry)  
XX  
XX DE Synthetic hirudin HVI-encoding sequence.  
XX  
XX KW anti-thrombin activity; thrombosis; blood clotting; ds.  
XX  
XX OS Synthetic.  
XX  
XX PN W09117250-A.  
XX  
XX PD 14-NOV-1991.  
XX  
XX PF 05-APR-1991; 91WO-EP00643.  
XX  
XX PR 10-MAY-1990; 90GB-0010552.  
XX  
XX PA (FARM ) FARMITALIA C ERBA SRL.  
XX  
XX PI Benatti L, Carminati P, Lansen J, Mazue G, Roncucci R;  
XX  
XX PI Sarmientosp, Scacheri E, De Taxis du Poet P;  
XX  
XX DR WPI; 1991-353771/48.  
XX  
XX PT Expression vectors encoding new and known hirudin(s) - or  
XX  
XX PT hirudin-like polypeptide(s), useful for treating thromboembolic  
XX  
XX PT events and thromboses etc.  
XX  
XX PS Example 1; Fig 1; 41pp; English.  
XX  
XX CC The coding sequence was designed on the basis of the E.coli  
XX  
XX CC preferred codons. A Bali restriction site was engineered close to  
XX  
XX CC the 5' end of the synthetic gene to allow insertion of the coding  
XX  
XX CC sequence in different expression vectors. The same synthetic gene  
XX  
XX CC can be used for expression in both bacterial and insect cells (for  
XX  
XX CC insect cells methods were developed which yielded secreted or  
XX  
XX CC cytoplasmic products. The HVI gene was constructed from four  
XX  
XX CC oligonucleotides and the double-stranded sequence has HindII and  
XX  
XX CC PstI sticky ends.  
XX  
XX SQ Sequence 185 BP; 52 A; 44 C; 51 G; 38 T; 0 other;

Query Match 76.3%; Score 82.4; DB 12; Length 185;

Best Local Similarity 89.0%; Pred. No. 7.1e-19; Indels 0; Gaps 0;

Matches 89; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 9 tattattgcaggtattctccgggatttctcaaaagtcgcgcgtgtgtgagactgcggt 68

Db 179 TACTATTGCAGGTATTCTTCCGGGATTCTTCGAAGTACACCGTCGTGTGACACTGCGGT 120

Qy 69 ttccggagtcaccttcgcagtaacgcgaattgatttttttcgc 108

Db 119 TTCCGCGGTACCTTCACCGGTAAACGACTGGTTTTTTCAC 80

## RESULT 9

AAQ10452/c  
ID AAQ10452 standard; DNA; 210 BP.

XX AC AAQ10452;  
XX  
XX DT 16-APR-1991 (first entry)  
XX  
XX DE HV-1 gene.

XX Hirudin; anticoagulant; fusion; porcine adenylate kinase.  
XX  
XX OS Synthetic.  
XX  
XX FH Key Location/Qualifiers  
XX  
XX FT CDS 7..219  
XX  
XX FT /\*tag= a  
XX  
XX PN EP412526-A.  
XX  
XX PD 13-FEB-1991.  
XX  
XX PF 08-AUG-1990; 90EP-0115226.  
XX  
XX PR 10-AUG-1989; 89JP-0207200.  
XX  
XX PA (NIHA ) NIPPON MINING KK.  
XX  
XX PI Misawa S, Matsuda H, Abe S;  
XX  
XX DR WPI; 1991-045687/07.  
XX  
XX DR P-PSDB; AAR10969.  
XX  
XX PT New expression vector for anticoagulant hirudin - transformed  
XX  
XX PT into Escherichia coli, is expressed in form of fused protein with  
XX  
XX PT porcine adenylate kinase  
XX  
XX PS Disclosure; Fig 1; 21pp; English.  
XX  
XX CC The gene was constructed from 8 fragments, and has sticky ends.  
XX  
XX CC The 5' terminal of the sense strand overhangs the 3' end of the  
XX  
XX CC complementary strand by 4 bases (EcoRI); the 5' terminal of the  
XX  
XX CC complementary strand overhangs the 3' end of the sense strand by  
XX  
XX CC -tcga- (HindIII). The synthetic gene was cloned into pUC18 to  
XX  
XX CC give pUCHV1. This was digested with NcoI and HindIII and ligated  
XX  
XX CC to construct pMAKHVI which was used to transform E.coli JMI09 for  
XX  
XX CC prodn. of a fusion protein of hirudin and adenylate kinase. When  
XX  
XX CC cleaved from the adenylate kinase the hirudin can be used to block  
XX  
XX CC blood coagulation through the inhibition of thrombin activity.  
XX  
XX CC See also AAQ10450-Q10454.  
XX  
XX SQ Sequence 210 BP; 60 A; 40 C; 55 G; 55 T; 0 other;

Query Match 71.9%; Score 77.6; DB 12; Length 210;

Best Local Similarity 86.0%; Pred. No. 3.3e-17; Indels 0; Gaps 0;

Matches 86; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 9 tattattgcaggtattctccgggatttctcaaaagtcgcgcgtgtgtgagactgcggt 68

Db 209 TACTACTGCAGGTATTCTTCCGGGATTCTTCGAATACACCATCGTTATGAGACTGCGGT 150

Qy 69 ttccggagtcaccttcgcagtaacgcgaattgatttttttcgc 108

Db 149 TTCCGGGTACCTTCACCGGTAAACGACTGGTTTTTTCAC 110

## RESULT 10

AAQ1295/c

ID AAQ1295 standard; DNA; 238 BP.

XX AC AAQ1295;  
XX  
XX DT 17-DEC-1990 (first entry)  
XX  
XX DE Sequence encoding hirudin HVI.  
XX  
XX KW Anticoagulant; thrombin inhibitor; yeast expression cassette; ss.  
XX  
XX OS Synthetic.  
XX



```

FT mat_peptide /*tag= a
FT 1..1242
FT /*tag= b
FT /*label= streptokinase
FT 1243..1254
FT /*tag= c
FT /*label= linker
FT /*note= "encodes factor Xa cleavage site"
FT 1255..1453
FT /*tag= d
FT /*label= hirudin HV-1
XX
XX W09109125-A.
XX
XX 27-JUN-1991.
XX
XX 07-DEC-1990; 90WO-GB01911.
XX
XX 07-DEC-1990; 90WO-GB01911.
XX 07-DEC-1989; 89GB-0027722.
XX
XX (BRBI-) BRIT BIO-TECHN LTD.
XX
XX Dawson KM, Hunter MG, Czaplewski LG;
XX
XX WPI: 1991-208151/28.
XX P-PSDB; AAR12522.
XX
XX Fusion protein cleavage by blood clotting enzyme - for prodn. of
XX fractions having greater antithrombotic activity for therapy and
XX prophylaxis.
XX
XX Disclosure; Page 98; 115pp; English.
XX
XX The sequence of the synthetic hirudin HV-1 gene was designed
XX based on the published amino acid sequence (Dodd J., et al FEBS
XX Letters 165 180 (1984)). The sequence of streptokinase was obtd.
XX from PCR amplified chromosomal DNA from S. equisimilis ATCC 10009
XX or ATCC 9642. The primers used for the PCR were based on the pub-
XX lished DNA sequence of S. equisimilis strain H46A (Malke, H., Roe,
XX B., and Ferretti, J.J., Gene 34 357-362 (1985)). The two
XX sequences were used to construct an expression vector in which the
XX streptokinase gene is linked to the hirudin gene via a linking
XX sequence encoding a cleavage site for factor Xa. The factor Xa is
XX present at the site of the target thrombus so the active agents are
XX released specifically at the place where clot formation is occurring.
XX See also AAQ12153-Q12156 and AAQ12158-Q12162.
XX
XX Sequence 1467 BP; 494 A; 317 C; 292 G; 364 T; 0 other;

Query Match 70.2%; Score 75.8; DB 12; Length 1467;
Best Local Similarity 83.5%; Pred. No. 2.3e-16;
Matches 86; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 6 cccattatgcaggattcttcgggattcttcgaagtcgcgcgtgttgagactgc 65
DB 1457 CCCTATTACTGCGAGATATTCTTCGGGATTTCTTCGAAATCTCCATCGTTGTGGACTGC 1398

QY 66 gggttcggagtagcttcgccagtaaacgcattgattttttcgc 108
DB 1397 GGCTTTGGGGTACCTTCACCAAGTGACACATGGTGTCTTTTCAC 1355

RESULT 13
AAQ54997/C
ID AAQ54997 standard; DNA; 567 BP.
XX
XX AAQ54997;
XX
XX 11-JUL-1994 (first entry)
XX
XX Desulphatohirudin gene expression/excretion cassette.

```

```

XX
XX Hirudin; HV-1; E. coli; codon usage; desulphatohirudin; HV01; 33Aasp;
XX 33 Asn; biological activity; thrombosis; thromboembolism; ss.
XX
XX Hirudo medicinalis.
XX
XX Key Location/Qualifiers
XX -35_signal 43..48
XX /*tag= a
XX -10_signal 68..73
XX /*tag= b
XX misc_signal 83
XX /*tag= c
XX /*note= "Transcription initiation site"
XX RBS 260..266
XX /*tag= d
XX /*note= "Shine-Delgarno sequence"
XX sig_peptide 269..352
XX /*tag= e
XX mat_peptide 353..550
XX /*tag= f
XX /*product= Hirudin
XX
XX EP576792-A.
XX
XX 05-JAN-1994.
XX
XX 13-APR-1993; 93EP-0105848.
XX
XX 09-APR-1992; 92HU-0001200.
XX (BIOG ) BIOGAL GYOGYSZERGYAR.
XX
XX Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;
XX Botond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;
XX Mate G, Molnari, Moravcsik I, Ott I, Patthy A;
XX Poelya K, Salat J, Tegdes A, Vincze A, Zilahi E;
XX
XX WPI: 1994-009153/02.
XX P-PSDB; AAR47490.
XX
XX Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using
XX E. coli, Saccharomyces and Streptomyces hosts, for increased
XX yields
XX
XX Disclosure; Page 48; 79pp; English.
XX
XX This sequence represents the an expression/secretion cassette for
XX the expression of hirudin HV-1. This sequence was used in the
XX production of desulphatohirudin HV01 33Aasp and desulphatohirudin HV-1
XX 33 Asn. The expressed peptides produced using this sequence have the
XX same biological activity as natural hirudin and can be used in the
XX treatment of thrombosis, thromboembolism, etc. Using naturalised
XX sequences such as this, large amounts of hirudin may be produced
XX stably, with the highest production level achieved being 140-180
XX mg/litre of culture.
XX
XX Sequence 567 BP; 162 A; 123 C; 134 G; 148 T; 0 other;

Query Match 69.4%; Score 75; DB 15; Length 567;
Best Local Similarity 81.3%; Pred. No. 3.4e-16;
Matches 87; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 2 gccgccttattatgcaggattcttcgggattcttcgaagtcgcgcgtgttgaga 61
DB 559 GCATGCTTACTATTGTAGGTATTTCTTCAGGAATTTCTTCGAAGTCACCATCTGTTGGGA 500

QY 62 ctgcgggttcggagtagcttcgccagtaaacgcattgattttttcgc 108
DB 499 TTGTGGCTTTGGAGTACCTTCGCCAGTGACACATTGATTTTTTTCAC 453

```

RESULT 14  
 AAQ54995/c  
 ID AAQ54995 standard; DNA; 219 BP.  
 XX AC AAQ54995;  
 XX AC AAQ54995;  
 XX AC AAQ54995;  
 DT 11-JUL-1994 (first entry)  
 XX DE Desulphatohirudin gene with E. coli-Saccharomyces codon usage.  
 XX DE Hirudin; HV-1; E. coli; codon usage; desulphatohirudin; HV01; 33Asp;  
 KW 33 Asp; biological activity; thrombosis; thromboembolism; ds.  
 XX OS Hirudo medicinalis.  
 XX OS  
 FH Key Location/Qualifiers  
 FT CDS 14..208  
 FT /\*tag= a  
 FT /product= Hirudin\_HV-1  
 XX PN EP576792-A.  
 XX PN  
 XX PD 05-JAN-1994.  
 XX PD  
 XX PF 13-APR-1993; 93EP-0105848.  
 XX PF  
 XX PR 09-APR-1992; 92HU-0001200.  
 XX PR  
 XX PA (BIOG ) BIOGAL GYGYSZERGYAR.  
 XX PA  
 XX PI Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;  
 PI Bocond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;  
 PI Mace G, Molnari, Moravcsik I, Ott I, Patthy A;  
 PI Poelya K, Salat J, Tegdes A, Vincze A, Zillahi E;  
 XX  
 DR WPI: 1994-009153/02.  
 DR P-PSDB; AAR47488.  
 XX  
 PT Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using  
 PT E. coli, Saccharomyces and Streptomyces hosts, for increased  
 PT yields  
 XX  
 PS Disclosure; Page 44; 79pp; English.  
 XX  
 CC This sequence represents the hirudin HV-1 gene designed on the  
 CC basis on E. coli codon usage. This sequence was used in the  
 CC production of desulphatohirudin HV01 33Asp and desulphatohirudin HV-1  
 CC 33 Asp. The expressed peptides produced using this sequence have the  
 CC same biological activity as natural hirudin and can be used in the  
 CC treatment of thrombosis, thromboembolism, etc. Using naturalised  
 CC sequences such as this, large amounts of hirudin may be produced  
 CC stably, with the highest production level achieved being 140-180  
 CC mg/litre of culture.  
 XX  
 SQ Sequence 219 BP; 63 A; 46 C; 51 G; 58 T; 1 other;

Query Match 69.3%; Score 74.8; DB 15; Length 219;  
 Best Local Similarity 83.3%; Pred. No. 3.1e-16;  
 Matches 85; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 7 cctattattgcaggattcttcctcgggatttcttcaaaagtcgcccgtctgttgagactgog 66  
 DB 215 CTATTATTGTAGGTATTCTTCAGGAATTCCTCGAAGTCACCATCGTGTGGGATTGTG 156  
 QY 67 gtttcgagtagctcttcgagtagcagcattgatttttttcgc 108  
 DB 155 GCTTGGAGTACCTTCGCCAGTGACACATGATTTTTCAC 114

RESULT 15  
 AAQ12153/c  
 ID AAQ12153 standard; DNA; 201 BP.

XX AAQ12153;  
 XX AC  
 XX 17-SEP-1991 (first entry)  
 XX AC  
 XX Synthetic hirudin type HV-1 gene.  
 DE  
 XX Fusion protein; blood clotting; coagulation; fibrinolysis;  
 KW antithrombotic; thrombolysis; streptokinase; ds.  
 XX OS Synthetic.  
 XX OS  
 FH Key Location/Qualifiers  
 FT CDS 1..201  
 FT /\*tag= a  
 XX PN W09109125-A.  
 XX PN  
 XX PD 27-JUN-1991.  
 XX PD  
 XX PF 07-DEC-1990; 90WO-GB01911.  
 XX PF  
 XX PR 07-DEC-1990; 90WO-GB01911.  
 XX PR  
 XX PR 07-DEC-1989; 89GB-0027722.  
 XX PR  
 XX PA (BRBI-) BRIT BIO-TECHN LTD.  
 XX PA  
 XX PI Dawson KM, Hunter MG, Czaplowski LG;  
 XX PI  
 XX WPI: 1991-208151/28.  
 XX P-PSDB; AAR12887.  
 DR  
 DR Fusion protein cleavage by blood clotting enzyme - for prodn. of  
 PT fractions having greater antithrombotic activity for therapy and  
 PT prophylaxis  
 PT  
 XX  
 PS Disclosure; Page 68; 115pp; English.  
 XX  
 CC The sequence was designed based on the published amino acid se-  
 CC quence (Dodt J., et al FEBS Letters 165 180 (1984)). Unique re-  
 CC striction sites were incorporated to facilitate subsequent genetic  
 CC manipulation. The codons selected were those favourable for S.  
 CC cerevisiae or E. coli. The sequence was divided into 12 oligomers  
 CC which were synthesised and then annealed. The ligation prod. was  
 CC ligated to HindIII and EcoRI treated pUC19 plasmid DNA and the  
 CC resulting vector used to transform E. coli K12 HW87. Plasmid pUC19  
 CC HV-1 was isolated from transformants and inserted into plasmid  
 CC pSW6, a shuttle vector, for expression. The gene can be used to  
 CC construct expression vectors in which the hirudin gene is linked to  
 CC a second gene encoding e.g. another hirudin protein, streptokinase  
 CC or a streptokinase-like protein, via a linking peptide. This pep-  
 CC tide link contains a cleavage site for e.g. factor X or thrombin  
 CC which can be cleaved, releasing the individual proteins which have  
 CC antithrombotic activity. The enzymes which cleave the fusion pro-  
 CC tein are present at the site of the target thrombus so the active  
 CC agents are released specifically at the place where clot formation  
 CC is occurring.  
 CC See also AAQ12154-Q12156, AAQ12158-Q12162 and AAQ12490.  
 XX  
 SQ Sequence 201 BP; 59 A; 43 C; 52 G; 47 T; 0 other;

Query Match 68.3%; Score 73.8; DB 12; Length 201;  
 Best Local Similarity 83.2%; Pred. No. 6.7e-16;  
 Matches 84; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 8 ctattattgcaggattcttcctcgggatttcttcaaaagtcgcccgtctgttgagactgog 67  
 DB 201 CTATTACTCAGATATCTCTTGGGATTTCTCGAATCTCCATCGTGTGGGACTGGG 142  
 QY 68 ttctcgagtagctcttcgagtagcagcattgatttttttcgc 108  
 DB 141 CTTTGGGGTACCTTCACAGTGCACACATGTTCTTTTCAC 101

Search completed: July 15, 2002, 22:44:49  
Job time: 10062 sec





Qy 64 ggggtttcggagtaccttcgccagtaacgcattgatttttttcgc 108  
 Db 62 GGGGTTTCGGGTACCTTCGCCACTAAGCACCTGGTGTGTTTTCGCG 106

## RESULT 2

US-08-262-384A-8  
; Sequence 8, Application US/08262384A  
; Patent No. 5624822

/ GENERAL INFORMATION:  
 / APPLICANT: Koerwer, Wolfgang  
 / TITLE OF INVENTION: The Preparation of Hirudin  
 / NUMBER OF SEQUENCES: 10  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Kell & Weinlauf  
 / STREET: 1101 Connecticut Avenue  
 / CITY: Washington  
 / STATE: D.C.  
 / COUNTRY: USA  
 /

[illegible]

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994

CLASSIFICATION:	530
CLASSIFICATION:	C 12 N 15/62
CLASSIFICATION:	C 12 N 15/31
CLASSIFICATION:	C 07 K 7/10
CLASSIFICATION:	C 12 P 21/02

;  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/EP90/02084  
 ; FILING DATE: 04-DEC-1990  
 ; APPLICATION NUMBER: US 07861820  
 ;

```

; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

US-08-262-384A-8

Query Match 88.3%; Score 95.4; DB 1; Length 235;  
Best Local Similarity 94.3%; pred. No. 3.6e-24;  
Matches 99; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

4 cgccctatttgaggattcttccgggatttcttcaagtcgcgcgtcggtgtgagact 63  
|||  
2 CGACCTATTACTGCAGGTATCTTCGCGGATTCCTCGAAGTCGCCGTCGTTCCTGAGACT 61  
|||

64 gcggttttcgagtagtaccttcgcagtaacgcattgatttttttcgc 108  
|||||  
62 GCGGTTTCGGGGTACCTTCGCCAGTACGCACATGGTGTTCCTTCGC 106  
|||||

### RESULT 3

US-08-262-384A-4/c  
Sequence 4, Application US/08262384A  
Patent No. 5624822

GENERAL INFORMATION:  
APPLICANT: Koerwer, Wolfgang  
TITLE OF INVENTION: The Preparation of Hirudin  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.

CITY: Washi  
STATE: D.C.

COUNTRY: USA  
ZIP: 20036

REF. 20030  
COMPUTER READABLE FORM:

**MEDIUM TYPE:** Diskette, 3.5 inch, 1.44 Mb storage  
**COMPUTER:** IBM AT-compatible, 80486 processor  
**OPERATING SYSTEM:** MS-DOS version 6.0  
**SOFTWARE:** WordPerfect version 5.1

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994

CLASSIFICATION:	530
CLASSIFICATION:	C 12 N 15/62
CLASSIFICATION:	C 12 N 15/31
CLASSIFICATION:	C 07 K 7/10
CLASSIFICATION:	C 12 P 21/02

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990

APPLICATION NUMBER: US 07861820

FILING DATE: 18-JUN-1992  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 106 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

-262-384A-4

Query Match 87.0%; Score 94; DB 1; Length 106;  
Best Local Similarity 95.1%; Pred. No. 8.5e-24;  
Matches 97; Conservative 0; Mismatches 5; Indels

Qy 7 cctattatgcaggattcttccgggatttcttcaaaagtcgccgctgatttgagactgog 66  
 |||||  
 Db 106 CCTATTACATGCAGGATTCTTCGGGATTTCTTCAAGTCGCCCGCTTGTGASACTGCG 47

Qy 67 gttcggagtagcttcgccagtaacgcattgatttttttcgc 108  
|||||  
Db 46 GTTTCGGGGGTACCTTCGCCAGTAAACGCATGCTTTTTCGC 5

## RESULT

US-08-262-384A-7/c  
; Sequence 7, Application US/08262384A  
; Patent No. 5624822.

```

; Patent No. 5024022
;
; GENERAL INFORMATION:
;
; APPLICANT: Koerwer, Wolfgang
;
; TITLE OF INVENTION: The Preparation of Hirudin
;
; NUMBER OF SEQUENCES: 10
;
; CORRESPONDENCE ADDRESS:

```

;  
 ;  
 ; ADDRESS: Keil & Weinkauff  
 ; STREET: 1101 Connecticut Avenue  
 ; CITY: Washington

STATE: D.C.  
COUNTRY: USA

ZIP: 20036

```

;
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
;
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
;
```

; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/262,384A  
 ; FILING DATE: 20-JUN-1994

;	CLASSIFICATION:	530	
;	CLASSIFICATION:	C 12 N 15/62	
;	CLASSIFICATION:	C 12 N 15/31	
;	CLASSIFICATION:	C 07 K 7/10	
;	CLASSIFICATION:	C 12 P 21/02	

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/EP90/02084  
 ; FILING DATE: 04-DEC-1990



**Qy**      8 ctattatgcagggtattcttcggttctcgaagttctccggcggttgttgagactgcgg 67  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Dp**    726 CTATTACTCAGGGTATCTTCGCGGAATTCTCCGAAGTCGCCGTTCGCAGACATCGCG 667

```

RESULT      7
US-08-186-222-3/c
: Sequence 3, Application US/08186222
: Patent No. 559007
: GENERAL INFORMATION:
: APPLICANT: Suri, Bruno
: APPLICANT: Schmitz, Albert
: TITLE OF INVENTION: Bacterial V
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corpora
: STREET: 7 Skylene Drive

```

```

; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186.222
; FILING DATE:
;
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672.205
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GB 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, Joann
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (recombinant)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Hybrid gene: MSP signal (base pair 1 to 81)/
; INDIVIDUAL ISOLATE: desulfatohirudin (base pair 82 to 279)
; IMMEDIATE SOURCE:
; CLONE: pUCRS/pML310
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..276
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 82..276
;
; US-08-186-222-3
;
; Query Match 78.0%; Score 84.2; DB 1; Length 279;
; Best Local Similarity 91.8%; Pred. No. 2.7e-20;
; Matches 89; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
;
; QY 12 tattgcaggtattcttcgaggtattcttcaaaagtcgcgcgtgtgtgagactcggttc 71
; Db 278 TACTGCAGGTATCTTCGGGATTTCTCGAAGTCACCGTGTGTGTGAGACTGCGGTTTC 219
;
; QY 72 ggaatcaccttcgcagtaacgcattgatttttcgc 108
; Db 218 GGGGTACCTTCGCCGGTAACCACTGGTTTTTTTCAC 182
;
; RESULT 8
; Patent No. 5180668
; APPLICANT: CRAUSE, PETER; HABERMANN, PAUL; TRIPIER, DOMINIQUE
; TITLE OF INVENTION: HIRUDIN DERIVATIVE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295.422
; FILING DATE: 10-JAN-1989
; SEQ ID NO: 10:
; LENGTH: 212
;
; 5180668-10
;
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186.222
; FILING DATE:
;
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672.205
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GB 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, Joann
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (recombinant)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Hybrid gene: MSP signal (base pair 1 to 81)/
; INDIVIDUAL ISOLATE: desulfatohirudin (base pair 82 to 279)
; IMMEDIATE SOURCE:
; CLONE: pUCRS/pML310
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..276
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 82..276
;
; US-08-186-222-3
;
; Query Match 78.0%; Score 84.2; DB 1; Length 279;
; Best Local Similarity 91.8%; Pred. No. 2.7e-20;
; Matches 89; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
;
; QY 12 tattgcaggtattcttcgaggtattcttcaaaagtcgcgcgtgtgtgagactcggttc 71
; Db 278 TACTGCAGGTATCTTCGGGATTTCTCGAAGTCACCGTGTGTGTGAGACTGCGGTTTC 219
;
; QY 72 ggaatcaccttcgcagtaacgcattgatttttcgc 108
; Db 218 GGGGTACCTTCGCCGGTAACCACTGGTTTTTTTCAC 182
;
; RESULT 8
; Patent No. 5180668
; APPLICANT: CRAUSE, PETER; HABERMANN, PAUL; TRIPIER, DOMINIQUE
; TITLE OF INVENTION: HIRUDIN DERIVATIVE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295.422
; FILING DATE: 10-JAN-1989
; SEQ ID NO: 10:
; LENGTH: 212
;
; 5180668-10
;
; Query Match 77.0%; Score 83.2; DB 6; Length 212;
; Best Local Similarity 87.5%; Pred. No. 5.5e-20;
; Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
;
; QY 5 gccctattattgcaggtattcttcgaggtattcttcaaaagtcgcgcgtgtgtgagactg 64
; Db 209 GCTCTATTACTGAAGGTATTTCCTCAGGGATCTCTTCGAAGTCGCGTGTATGAGACTG 150
;
; QY 65 cggtttcgaggtacaccttcgcagtaacgcattgatttttttcgc 108
; Db 149 CGGTTTCGGGTACCTTCGCCAGTACGCACTGGTCTTTTTCAC 106
;
; RESULT 9
; US-07-982-064-8/c
; Sequence 8, Application US/07982064
; Patent No. 5919895
; GENERAL INFORMATION:
; APPLICANT: Schmid, G.; Habermann, P.
; TITLE OF INVENTION: Secretion of Hirudin Derivatives
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Collard, Roe & Galgano, P.C.
; STREET: 1077 No. 5919895thern Boulevard
; CITY: Roslyn
; STATE: New York
; COUNTRY: USA
; ZIP: 11576
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/982,064
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/672,791
; FILING DATE: 21-MAR-1991
; APPLICATION NUMBER: GR 40 09 268.2
; FILING DATE: 22 MAR 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Collard, Allison C.
; REGISTRATION NUMBER: 22,532
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W3
; NAME: Galgano, Thomas M.
; REGISTRATION NUMBER: 27,638
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W3
; NAME: Freedman, Edward R.
; REGISTRATION NUMBER: 26,048
; REFERENCE/DOCKET NUMBER: SCHMID ET AL-W3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-365-9802
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 227 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other DNA
;
; US-07-982-064-8
;
; Query Match 77.0%; Score 83.2; DB 2; Length 227;
; Best Local Similarity 87.5%; Pred. No. 5.7e-20;
; Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
;
; QY 5 gccctattattgcaggtattcttcgaggtattcttcaaaagtcgcgcgtgtgtgagactg 64
; Db 201 GCTCTATTACTGAAGGTATTTCCTCAGGGATCTCTTCGAAGTCGCGTGTATGAGACTG 142
;
; QY 65 cggtttcgaggtacaccttcgcagtaacgcattgatttttttcgc 108
```



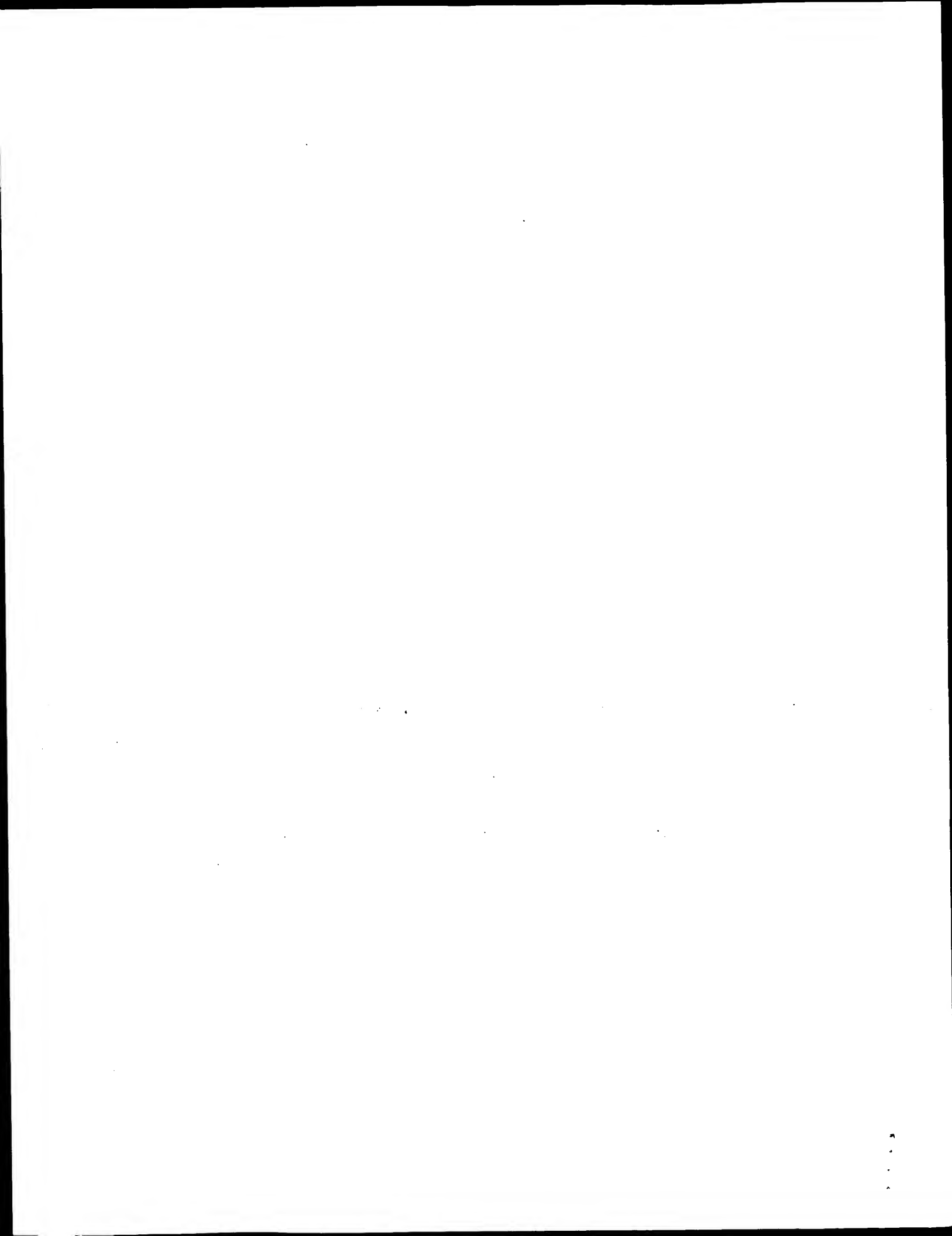


Tue Jul 16 08:24:09 2002

us-10-053-641-4.rni

---

Page 7



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:48:54 ; Search time 6165.88 Seconds  
(without alignments)  
236.409 Million cell updates/sec

Title: US-10-053-641-4  
Perfect score: 108  
Sequence: 1 ggcgcctattattgcagg.....aacgcattgatttttcgc 108

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*\*
  - 2: em\_esthum:\*\*
  - 3: em\_estin:\*\*
  - 4: em\_estmu:\*\*
  - 5: em\_estor:\*\*
  - 6: em\_estpi:\*\*
  - 7: em\_estro:\*\*
  - 8: em\_htc:\*\*
  - 9: gb\_estl:\*\*
  - 10: gb\_est2:\*\*
  - 11: gb\_htc:\*\*
  - 12: gb\_gss:\*\*
  - 13: em\_gss\_hum:\*\*
  - 14: em\_gss\_inv:\*\*
  - 15: em\_gss\_pln:\*\*
  - 16: em\_gss\_vrt:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	30.6	621	10	BG671298
2	33	30.6	644	10	BI107752
3	32.4	30.0	402	12	TA210G12Q
C 4	32.4	30.0	511	12	TA54D01P
5	29.8	27.6	678	10	BI798501
C 6	29.6	27.4	978	9	AL554921
C 7	29.6	27.4	1244	10	BG290977
8	29.4	27.2	387	12	BH214991
C 9	29.2	27.0	352	10	BM218166
10	29.2	27.0	626	9	AW536772
C 11	28.8	26.7	353	9	AW738920
12	28.8	26.7	432	12	AZ180488
C 13	28.8	26.7	495	10	BJ174087
C 14	28.8	26.7	551	10	BJ157738
C 15	28.8	26.7	577	10	BJ202386
C 16	28.8	26.7	578	10	BJ202320
C 17	28.8	26.7	579	10	BJ203448

C 18	28.8	26.7	580	10	BJ202387
19	28.8	26.7	601	10	BJ165686
C 20	28.6	26.5	906	12	AZ202441
21	28.6	26.5	1030	12	CNS034KT
C 22	28.4	26.3	200	10	228896
23	28.4	26.3	219	9	AA568870
24	28.4	26.3	232	9	BB586156
C 25	28.4	26.3	342	10	BF052721
C 26	28.4	26.3	390	9	AA282103
C 27	28.4	26.3	406	9	AA281638
28	28.4	26.3	435	10	H96697
C 29	28.4	26.3	457	9	AI685214
C 30	28.4	26.3	460	10	BF054388
C 31	28.4	26.3	478	9	AA938721
C 32	28.4	26.3	565	9	AV916968
C 33	28.4	26.3	567	10	BE499455
34	28.4	26.3	599	10	BI792095
35	28.4	26.3	699	12	BH547832
36	28.4	26.3	719	9	AV935444
37	28.4	26.3	851	10	BE704576
38	28.2	26.1	354	9	BE039897
39	28.2	26.1	396	9	AU082915
40	28.2	26.1	588	9	AU029435
41	28.2	26.1	609	9	AU097671
42	28.2	26.1	625	10	BI806759
43	28.2	26.1	627	9	AU086669
44	28.2	26.1	669	9	AU076129
45	28.2	26.1	708	9	AU097672

ALIGNMENTS

RESULT 1

BG671298  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

621 bp mRNA linear EST 30-APR-2001  
DRNBOD03 Rat DRG Library Rattus norvegicus cDNA clone DRNBOD03 5',  
mRNA sequence.  
BG671298  
BG671298.1 GI:13893397  
EST.  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 621)  
Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G.,  
Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and  
Zhang,X.  
Distinct gene expression profiles of rat dorsal root ganglion  
induced by peripheral nerve axotomy  
Unpublished (2001)  
Contact: Zhang Xu  
Laboratory of Sensory System  
Institute of Neuroscience  
320 Yue Yang Road, Shanghai 200031, P.R.China  
Tel: 86-21-64748700-121  
Fax: 86-21-64713446  
Email: xu.zhang@ion.ac.cn  
This clone is also available at Chinese National Human Genome  
Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,  
Pudong New Area, P.R.China. Please contact with Zhang Xu  
(xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)  
PCR Primers  
FORWARD: T3  
BACKWARD: T7  
Seq primer: T3  
POLYA=No.  
Location/Qualifiers  
1..621  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"





ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 978)
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequenage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.
FEATURES	Location/Qualifiers.

```

FEATURES
source
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .978
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSOD1087P21"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-Oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
320 a 223 c 235 g 190 t 10 others
BASE COUNT
ORIGIN
Query Match 27.4%; Score 29.6; DB 9; Length 978;
Best Local Similarity 62.5%; Pred. No.54;
Matches 45; Conservative 1; Mismatches 26; Indels 0; Gaps 0;

```

```

Qy 5 gccctattgcaggatttccgggatttcttccaaagtcgccgctgttgagactg 64
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 GGCCTCTGTTGCTGGATTTTCTCGAGTCTCTGCCAAGTGGVGTTCAGATACAGACTG 147
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 65 gcggttcggagt 76
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 CTGAGTTATAGT 135

RESULT 7
BG290977/c 1244 bp mRNA linear EST 21-FEB-2001
LOCUS 602386963F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516023 5',
DEFINITION mRNA sequence.
ACCESSION BG290977
VERSION BG290977.1 GI:13048483
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1244)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
CDNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M10406 row: m column: 16
High quality sequence stop: 653.
FEATURES
Location/Qualifiers
  1..1244
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4516023"
    /clone_lib="NIH_MGC_93"
    /tissue_type="transitional cell papilloma, cell line"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: bladder; Vector: pCMV-Sport6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
    Average insert size 1.7 kb. Library enriched for
    full-length clones and constructed by Life Technologies.
    Note: this is a NIH_MGC Library."
BASE COUNT 406 a 312 c 247 g 279 t
ORIGIN

Query Match 27.4%; Score 29.6; DB 10; Length 1244;
Best Local Similarity 61.8%; Pred. No. 56;
Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 17 cagggtatttccgggatttccaaagtcgccgctgttgagactgcggttcggagt 76
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 921 CCGGGGTCCTCTGTTGTTTCAATGTCGGCGTGTGAGAAATGCAGTGTAGGACC 862
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 77 accttcgcagtaacg 92
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 861 CATTTTGCACGAATG 846
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
BH214991 387 bp DNA linear GSS 08-NOV-2001
LOCUS 1006012F09.2EL_x1 1006 - RescueMu Grid G Zea mays genomic, DNA
DEFINITION sequence.
ACCESSION BH214991
VERSION BH214991.1 GI:16805681

```

```

KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 387)
AUTHORS Walbot, V.
TITLE Maize genomic sequences found using engineered RescueMu transposon
JOURNAL Unpublished (2001)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1006012 row: 38
Class: transposon-tagged.
FEATURES
Location/Qualifiers
  1..387
    /organism="Zea mays"
    /cultivar="mixed background W23/A188/B73"
    /db_xref="taxon:4577"
    /clone_lib="1006 - RescueMu Grid G"
    /dev_stage="adult"
    /tissue_type="leaf"
    /lab_host="DH10B"
    /note="Organ: leaf; Vector: RescueMu (engineered from
    pBluescript backbone); Site_1: BamHI; Site_2: BglII;
    RescueMu is a 4.9 kb, modified maize Mu transposon
    designed to allow plasmid rescue from total genomic DNA.
    Mu elements insert preferentially into transcription
    units. For more information on RescueMu, go to the web
    site 'www.zmdb.iastate.edu' and follow the links for
    'RescueMu.' Grid G was grown at Stanford in 2000. DNA was
    extracted from leaf punches, double digested using BamHI
    and BglII, and ligated to form circular plasmids. DH10B
    cells were transformed and then screened on LB plates with
    ampicillin."
BASE COUNT 101 a 98 c 94 g 93 t 1 others
ORIGIN

Query Match 27.2%; Score 29.4; DB 12; Length 387;
Best Local Similarity 56.8%; Pred. No. 54;
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 4 cggcctatttcgaggtattcttcgggatttcttccaaagtcgccgctgttgagact 63
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36 CCCCACCTTGATCCATGATAGCTTCAGCCATCTTTTAAATTCATCATCTTAGGCT 95
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 64 gcggttcggagtagcttcgccagtaacgcattga 98
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 GAGGCCCGGAGACGCTTCAGTAGGATCTGAGGA 130
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
BM218166/c 352 bp mRNA linear EST 31-JAN-2002
LOCUS BM218166
DEFINITION C0907B09-3 NIA Mouse 12.5-dpc Male Genital Ridge/Mesonephros cDNA
Library (Long) Mus musculus cDNA clone C0907B09 3', mRNA sequence.
ACCESSION BM218166
VERSION BM218166.1 GI:17777596
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 352)
AUTHORS Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Luo, A.

```



DEFINITION gbl6c05.v1 Moss EST library PPN Physcomitrella patens cDNA clone  
 REP\_SOURCE\_ID: PPN101009 5' similar to TR:Q9ZM57 Q9ZM57 F17L21.12.  
 ; mRNA sequence.  
 ACCESSION AW738920  
 VERSION AW738920.1 GI:7647937  
 KEYWORDS EST.  
 SOURCE Physcomitrella patens.  
 ORGANISM Physcomitrella patens.  
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Bryophyta:  
 Bryopsida: Funariidae: Funariales: Funariaceae: Physcomitrella.  
 1 (bases 1 to 353)  
 Quatrano, R., Bashardes, S., Cove, D., Cumming, A., Knight, C., Clifton  
 , S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood  
 , K., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,  
 Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,  
 Waterston, R., and Wilson, R.  
 Leeds/Wash U Moss EST Project  
 Unpublished (1999)  
 Contact: Ralph Quatrano  
 Leeds/Wash U Moss EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Libraries were constructed by Dr. Stavros Bashardes as part of the  
 Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and  
 Washington Univ. in St. Louis (USA) DNA sequencing by: Washington  
 University Genome Sequencing Center For information on obtaining a  
 clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)  
 Seq primer: -40RP from Gibco.  
 Location/Qualifiers  
 1..353  
 /organism="Physcomitrella patens"  
 /db\_xref="taxon:3218"  
 /clone="PEP\_SOURCE\_ID: PPN101009"  
 /clone\_lib="Moss EST library PPN"  
 /tissue\_type="protonemata: 7 day old tissue auxin treated"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
 XhoI; Construction of the cDNA library was carried out  
 using Stratagene 'Unizap - cDNA synthesis kit'. cDNA was  
 constructed using an oligo dt primer/linker that contains  
 a XhoI site within it. Following ds cDNA synthesis,  
 EcoRI adapters were ligated to the blunt ends and sample  
 was digested with XhoI. The result is cDNA with an EcoRI  
 sticky end on one side and a XhoI sticky end on the other.  
 This cDNA was ligated directionally in Unizap arms. The  
 vector is designed containing the pBluescript sequence as  
 well as lambda DNA and cDNA is cloned within this  
 pBluescript sequence. The vector was then packaged using  
 Gold gigapackaging extracts. Library was grown in XLBlue  
 MRF<sup>+</sup> cells and amplified. The library was excised by mass  
 excision using Stratagene 'Mass excision kit' that uses  
 exsist as a helper phage that releases the pBluescript  
 sequence and circularises it as single stranded plasmids  
 that are then packaged (by helper phage) and secreted out  
 of the host cell as phagemids. SOLR cells were transformed  
 with phagemids and the library was plated out on LB-amp  
 plates to select for transformants. Approximately 1,000  
 ,000 colonies were grown and recovered. The double  
 stranded plasmid library was recovered by using Quiaagen  
 Midi prep kit. 2 micro grams of each library were used to  
 transform DH10B cells by electroporation."

# FEATURES

source

1..353  
 /organism="Physcomitrella patens"  
 /db\_xref="taxon:3218"  
 /clone="PEP\_SOURCE\_ID: PPN101009"  
 /clone\_lib="Moss EST library PPN"  
 /tissue\_type="protonemata: 7 day old tissue auxin treated"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
 XhoI; Construction of the cDNA library was carried out  
 using Stratagene 'Unizap - cDNA synthesis kit'. cDNA was  
 constructed using an oligo dt primer/linker that contains  
 a XhoI site within it. Following ds cDNA synthesis,  
 EcoRI adapters were ligated to the blunt ends and sample  
 was digested with XhoI. The result is cDNA with an EcoRI  
 sticky end on one side and a XhoI sticky end on the other.  
 This cDNA was ligated directionally in Unizap arms. The  
 vector is designed containing the pBluescript sequence as  
 well as lambda DNA and cDNA is cloned within this  
 pBluescript sequence. The vector was then packaged using  
 Gold gigapackaging extracts. Library was grown in XLBlue  
 MRF<sup>+</sup> cells and amplified. The library was excised by mass  
 excision using Stratagene 'Mass excision kit' that uses  
 exsist as a helper phage that releases the pBluescript  
 sequence and circularises it as single stranded plasmids  
 that are then packaged (by helper phage) and secreted out  
 of the host cell as phagemids. SOLR cells were transformed  
 with phagemids and the library was plated out on LB-amp  
 plates to select for transformants. Approximately 1,000  
 ,000 colonies were grown and recovered. The double  
 stranded plasmid library was recovered by using Quiaagen  
 Midi prep kit. 2 micro grams of each library were used to  
 transform DH10B cells by electroporation."

# BASE COUNT

ORIGIN

76 a 77 c 100 g 100 t  
 Query Match 26.7%; Score 28.8; DB 9; Length 353;  
 Best Local Similarity 60.0%; Pred. No. 82;  
 Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
 QY 26 ttccgggttttttcaagtcgcgtctgttgagactcgcgttttcgagtagtacctgcc 85

Db 83 TACTGTTCTTCTCTCTTAATCGTCTCTGGGAGACTGCCCTCTCTGTGCAGTTCTTC 24  
 QY 86 agtaacgcattgattttt 105  
 Db 23 TGGAACTTAGCGCTCTTCT 4  
 RESULT 12  
 AZ180488  
 LOCUS  
 DEFINITION  
 SP\_0168\_B1\_E03\_SP6E Strongylocentrotus purpuratus, purple sea  
 urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
 genomic clone Plate-168 Col-5 Row-J, DNA sequence.  
 ACCESSION AZ180488  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Strongylocentrotus purpuratus.  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Euechinoidea; Echinacea; Echinoida;  
 Strongylocentrotidae; Strongylocentrotus.  
 1 (bases 1 to 432)  
 Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,  
 Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray  
 , G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and  
 Hood, L.  
 A sea urchin genome project: Sequence scan, virtual map, and  
 additional resources  
 Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
 20402566  
 Contact: Cameron, RA, Davidson, EH, Hood, L  
 Division of Biology 156-29  
 California Institute of Technology  
 Pasadena California 91125, USA  
 Tel: (626) 395-8421  
 Fax: (626) 793-3047  
 Email: acameron@caltech.edu  
 Plate: 168 row: J column: 5  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 432.  
 Location/Qualifiers  
 1..432  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone="Plate-168 Col-5 Row-J"  
 /clone\_lib="Strongylocentrotus purpuratus, purple sea  
 urchin, sperm genomic BAC library"  
 /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli  
 DH10B"

BASE COUNT 130 a 83 c 76 g 142 t 1 others  
 ORIGIN  
 Query Match 26.7%; Score 28.8; DB 12; Length 432;  
 Best Local Similarity 62.5%; Pred. No. 84;  
 Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
 QY 5 gccctatttcaggtatttcctccgggatttcctcaagtcgcgtctgtgagactg 64  
 Db 195 GCATTATCATTCACAAATTCCTCCGAGATTTCCTCGGAGATGACTATGTGAGAGAATA 254  
 QY 65 cgggttcggagt 76  
 Db 255 CATTCCTGGGACT 266

RESULT 13  
 BJ174087/c  
 LOCUS  
 DEFINITION  
 495 bp mRNA linear EST 24-JAN-2002  
 BJ174087 normalized full length cDNA library, chloronemata,  
 caulonemata and malformed buds Physcomitrella patens subsp. patens



Search completed: July 15, 2002, 21:49:01  
Job time: 18892 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:28:44 ; Search time 2368.24 Seconds  
(without alignments)  
291.599 Million cell updates/sec

Title: US-10-053-641-5  
Perfect score: 33  
Sequence: 1 tcggatcccttattggtgttacactgactgc 33

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
--------	-------	-------	-------	--------	----	----	-------------

1	25.2	76.4	226	12	SYNHIR	M26762 Synthetic h
2	23.2	70.3	150214	9	AC011291	AC011291 Homo sapi
c 3	23.2	70.3	188732	2	AC024050	AC024050 Homo sapi
4	23.2	70.3	188237	2	AP000906	AP000906 Homo sapi
5	23.2	70.3	196695	2	AC078936	AC078936 Homo sapi
c 6	23.2	70.3	269711	2	AP000408	AP000408 Homo sapi
7	22.4	67.9	33	6	A00633	A00633 Nucleotide
8	22.4	67.9	220	12	SYNHIRMED	M14964 Synthetic h
9	22.4	67.9	224	6	A00630	A00630 Artificial
c 10	22.4	67.9	224	6	A00631	A00631 Artificial
11	22.4	67.9	224	6	A04277	A04277 Artificial
c 12	22.4	67.9	224	6	A04278	A04278 Artificial
13	22.4	67.9	224	6	A07345	A07345 Synthetic D
c 14	22.4	67.9	224	6	A07346	A07346 Synthetic D
c 15	22	66.7	34	6	A00634	A00634 Nucleotide
16	22	66.7	93	6	I63540	I63540 Sequence 3
17	21.8	66.1	80291	3	AC096493	AC096493 Rattus no
18	21.4	64.8	35875	3	CE2C15	Z93396 Caenorhabdi
c 19	21.4	64.8	155550	2	AC020227	AC020227 Drosophi
c 20	21.4	64.8	162387	3	AC008233	AC008233 Drosophi
21	21.4	64.8	171952	10	AC002324	AC002324 Mus muscu
22	21.4	64.8	192550	2	AC025584	AC025584 Mus muscu
c 23	21.4	64.8	217238	2	AL591131	AL591131 Mus muscu
24	21.4	64.8	232750	2	AL591177	AL591177 Mus muscu
c 25	21.4	64.8	241429	3	AE003608	AE003608 Drosophi
c 26	21.2	64.2	84478	9	AL133462	AL133462 Human DNA
c 27	21	63.6	106	6	I41186	I41186 Sequence 6
28	21	63.6	129	6	I41183	I41183 Sequence 3
29	21	63.6	235	6	I41187	I41187 Sequence 7
c 30	21	63.6	235	6	I41188	I41188 Sequence 8
31	21	63.6	726	6	I41189	I41189 Sequence 9
c 32	21	63.6	726	6	I41190	I41190 Sequence 10
c 33	21	63.6	4809	6	AX200939	AX200939 Sequence
c 34	21	63.6	4809	6	AX267595	AX267595 Sequence
c 35	21	63.6	72273	2	AC090859	AC090859 Homo sapi
c 36	21	63.6	143420	9	AL161422	AL161422 Human DNA
37	20.8	63.0	58	6	A34231	A34231 Synthetic h
38	20.8	63.0	109	6	A34234	A34234 Synthetic h
c 39	20.8	63.0	109	6	A34235	A34235 Synthetic h
40	20.8	63.0	217	6	A03695	A03695 Nucleotide
c 41	20.8	63.0	217	6	A03696	A03696 Nucleotide
42	20.8	63.0	217	6	A34238	A34238 Synthetic d
c 43	20.8	63.0	217	6	A34239	A34239 Synthetic d
44	20.8	63.0	217	6	E00657	E00657 DNA encodin
45	20.8	63.0	168111	9	HS525L6	AL023807 Human DNA

ALIGNMENTS

RESULT	1	SYNHIR	SYNHIR	226 bp	DNA	linear	SYN 27-APR-1993
LOCUS	SYNHIR	Synthetic hirudin gene, complete cds.					
DEFINITION	1	(bases 1 to 226)					
ACCESSION	M26762						
VERSION	M26762.1	GI:208478					
KEYWORDS	hirudin; proteinase inhibitor; thrombin inhibitor.						
SOURCE	Synthetic DNA.						
ORGANISM	artificial sequence.						
REFERENCE	1	(bases 1 to 226)					
AUTHORS	Bergmann,C., Dodt,J., Koehler,S., Fink,E. and Gassen,H.G.						
TITLE	Chemical synthesis and expression of a gene coding for hirudin, the						
JOURNAL	thrombin-specific inhibitor from the leech Hirudo medicinalis						
MEDLINE	Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)						
FEATURES	87026239						
source	Location/Qualifiers						
	1..226						
	/organism="synthetic construct"						
	/db_xref="taxon:32630"						
	22..222						
	/note="hirudin"						
	/codon_start=1						
CDS							

```

/transl_table=11
/protein_id="AAA27772.1"
/db_xref="GI:208479"
/translations="MVYTDCTESGNLCLCEGSGVCGGNKCLILGSDGEKNCQVTGE
GTPKQSHNDGDFEIPPEYLQ"
BASE COUNT      63 a   56 c   61 g   46 t
ORIGIN

```

```

Query Match      76.4%; Score 25.2; DB 12; Length 226;
Best Local Similarity 90.0%; Pred. No. 1.4;
Matches 27; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 4 ggatccttatgtgtttacactgaatgc 33
| | ||||| ||||| ||||| ||||| |||||
Db 13 GTAAGCTTTATGGTTGTTTACACTGACTGC 42

```

```

RESULT 2
AC011291 AC011291 150214 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-67G7 from 2, complete sequence.
DEFINITION
AC011291
AC011291.8 GI:14589665
VERSION
HTG.
SOURCE
human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 150214)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792

```

```

REFERENCE 2 (bases 1 to 150214)
AUTHORS Mulvaney, E., Boyer, E. and Kozlowski, A.
TITLE The sequence of Homo sapiens BAC clone RP11-67G7
JOURNAL Unpublished (2001)

```

```

REFERENCE 3 (bases 1 to 150214)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

REFERENCE 4 (bases 1 to 150214)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

REFERENCE 5 (bases 1 to 150214)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

```

```

COMMENT On Jul 3, 2001 this sequence version replaced gi:14018119.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
-----
Summary Statistics
-----
Center project name: H_NH0067G07

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >=

30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

#### VECTOR: pBAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-693G13, 2000 bp overlap; the clone sequenced to the right is RP11-434M17, 2000 bp overlap. Actual end of this clone is at base position 14557 of RP11-434M17.

The sequence of RP11-67G7 from 94121 to 94182 is from a PCR product of clone DNA.

#### FEATURES

source	Location/Qualifiers
1..150214	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-67G7"
831..1416	/clone.lib="RPCI-11"
	/rpt_family="L2"
1832..2326	/note="match to EST AA885327 (NID:g2994404) al60g03.sl"
1833..2321	/note="match to EST BG196474 (NID:g13718161)"
1837..2120	/note="match to EST H96169 (NID:g1109311) yt98b10.sl"
1849..2114	/note="match to EST BG206827 (NID:g13728514)"
2161..2386	/note="similar to Homo sapiens EST AA757626 (NID:g2805489)
2942a02.sl"	
2325..2345	/rpt_family="(TG)n"
3080..3662	/note="match to EST H23520 (NID:g8922215) ym53d01.r1"
3167..3435	/note="match to EST R50919 (NID:g812821) yg70h03.r1"
3399..3515	/rpt_family="CT-rich"
3537..4947	/rpt_family="L1"
4948..5121	/rpt_family="L1"
5873..6004	/rpt_family="MIR"
6006..6096	/rpt_family="CRL"
6100..6137	/rpt_family="(TTTTTA)n"
6109..6432	/rpt_family="Alu"
7267..7327	/rpt_family="L1"
7475..7506	/rpt_family="AT-rich"



repeat\_region 8933. .8984  
/rpt\_family="T)n"  
repeat\_region 11892. .11965  
/rpt\_family="MIR"  
repeat\_region 12710. .12914  
/rpt\_family="L2"  
repeat\_region 12950. .13072  
/rpt\_family="MIR"  
repeat\_region 13265. .13302  
/rpt\_family="TATG)n"  
misc\_feature 13599. .14264  
/note="match to EST AL557921 (NID:g12901993)"  
misc\_feature 13600. .13852  
/note="match to EST BG184202 (NID:g13705889)"  
misc\_feature 13656. .14143  
/note="match to EST H94889 (NID:g1102522) yu57g08.sl"  
misc\_feature 13863. .14766  
/note="match to EST AL579862 (NID:g12945319)"  
repeat\_region 14186. .14213  
/rpt\_family="AT-rich"  
repeat\_region 14211. .14312  
/rpt\_family="MIR"  
misc\_feature 14782. .14785  
/note="match to EST AL579862 (NID:g12945319)"  
repeat\_region 14806. .14827  
/rpt\_family="AT-rich"  
misc\_feature 15159. .15479  
/note="match to EST AA774008 (NID:g2825897) ab67f02.rl"  
misc\_feature 15302. .15776  
/note="match to EST A1277366 (NID:g3899634) qm54b12.xl"  
repeat\_region 15961. .15966  
/rpt\_family="(TA)n"  
repeat\_region 16083. .16277  
/rpt\_family="Alu"  
repeat\_region 16368. .16473  
/rpt\_family="Alu"  
repeat\_region 16445. .16464  
/rpt\_family="(CAAAA)n"  
repeat\_region 16598. .17048  
/rpt\_family="(TG)n"  
repeat\_region 17257. .17330  
/rpt\_family="(TA)n"  
repeat\_region 17570. .17724  
/rpt\_family="MIR"  
repeat\_region 18450. .18505  
/rpt\_family="L2"  
repeat\_region 18852. .19013  
/rpt\_family="MER1\_type"  
repeat\_region 19224. .19351  
/rpt\_family="MER1\_type"  
repeat\_region 19311. .19418  
/rpt\_family="GA-rich"  
repeat\_region 19443. .19510  
/rpt\_family="MER1\_type"  
repeat\_region 19851. .20074  
/rpt\_family="MIR"  
repeat\_region 20084. .20118  
/rpt\_family="Mariner"  
repeat\_region 20201. .20258  
/rpt\_family="MIR"  
repeat\_region 20305. .20409  
/rpt\_family="L2"  
misc\_feature 20666. .20679  
/note="match to EST BG214571 (NID:g13740592)"  
repeat\_region 20666. .20959  
/rpt\_family="Alu"  
repeat\_region 21013. .21049  
/rpt\_family="(CA)n"  
repeat\_region 21508. .21544  
/rpt\_family="(TG)n"  
repeat\_region 23468. .23634  
/rpt\_family="MIR"  
repeat\_region 24535. .24643

Query Match 70.3%; Score 23.2; DB 9; Length 150214;  
Best Local Similarity 89.3%; Pred. No. 7;  
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
/rpt\_family="MIR"  
QY 4 ggatcctttatggtttttacactgact 31  
||||| | |||||||||  
Db 6647 GGATCCTGTTAGTTGTTTACACTGACT 6674  
RESULT 3  
AC024050/c  
LOCUS  
DEFINITION Homo sapiens chromosome 11 clone RP11-569A20, WORKING DRAFT  
AC024050  
AC024050.6 GI:7637374  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 168732)  
Waterston,R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 168732)  
Waterston,R.H.  
Direct Submission  
Submitted (20-FEB-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Apr 22, 2000 this sequence version replaced gi:7579880.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H\_NH0569A20  
----- Summary Statistics -----  
Sequencing vector: pLM3; 77%  
Chemistry: Dye-terminator Big Dye; 23% of reads  
Chemistry: Dye-terminator Big Dye; 77% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 166423 bases at least Q40  
Consensus quality: 166920 bases at least Q30  
Consensus quality: 167253 bases at least Q20  
Insert size: 166000; agarose-fp  
Quality coverage: 168332; sum-of-contigs  
Quality coverage: 7.51 in Q20 bases; agarose-fp  
Quality coverage: 7.44 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1680: contig of 1680 bp in length  
\* 1681 1780: gap of unknown length  
\* 1781 9168: contig of 7388 bp in length  
\* 9169 9268: gap of unknown length  
\* 9269 26141: contig of 16873 bp in length  
\* 26142 26241: gap of unknown length  
\* 26242 87019: contig of 60778 bp in length  
\* 87020 87119: gap of unknown length  
\* 87120 168732: contig of 81613 bp in length.  
Location/Qualifiers

FEATURES

Tue Jul 16 08:24:10 2002

```

source
1. .168732
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-569A20"

BASE COUNT 54795 a 30636 c 29985 g 52905 t 411 others
ORIGIN

Query Match 70.3%; Score 23.2; DB 2; Length 168732;
Best Local Similarity 89.3%; Pred. No. 7;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gatcccttatggtgtttacactgactg 32
|||||
Db 73666 GATCCTTTATGATGTTTCTCTGACTG 73639
|||||

RESULT 4
AP000906 188237 bp DNA linear HTG 30-MAY-2000
LOCUS Homo sapiens chromosome 11 clone RP11-70717 map 11q22, WORKING
DEFINITION DRAFT SEQUENCE, 43 unordered pieces.
ACCESSION AP000906.3 GI:8119048
VERSION HTG; HTGS-PHASE1; HTGS-DRAFT.
KEYWORDS Homo sapiens DNA, clone:RP11-70717.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 188237)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 188,237 genomic DNA of 11q22
Published only in Database (1999) In press
2 (bases 1 to 188237)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997724.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: RP11-70717
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator Et-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167827 bases at least Q40
Consensus quality: 177043 bases at least Q30
Consensus quality: 181388 bases at least Q20
Insert size: 184037; sum-of-contigs
Quality coverage: 4.22x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
43 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 14984 contig of 14984 bp in length
15085 25511 contig of 10427 bp in length

35208 contig of 9597 bp in length
45403 contig of 10095 bp in length
54399 contig of 8896 bp in length
63206 contig of 8707 bp in length
70841 contig of 7535 bp in length
79969 contig of 9028 bp in length
88468 contig of 8399 bp in length
94989 contig of 6421 bp in length
100503 contig of 5314 bp in length
110335 contig of 4893 bp in length
115335 contig of 4900 bp in length
119536 contig of 4101 bp in length
124058 contig of 4422 bp in length
128684 contig of 4526 bp in length
132388 contig of 3604 bp in length
136139 contig of 3651 bp in length
139193 contig of 2954 bp in length
142207 contig of 2914 bp in length
146335 contig of 4028 bp in length
149853 contig of 3418 bp in length
152741 contig of 2788 bp in length
154497 contig of 1656 bp in length
156939 contig of 2342 bp in length
157559 contig of 520 bp in length
159511 contig of 1852 bp in length
162263 contig of 2852 bp in length
164148 contig of 1785 bp in length
166170 contig of 1922 bp in length
168674 contig of 2394 bp in length
170667 contig of 1913 bp in length
173092 contig of 2315 bp in length
175018 contig of 1826 bp in length
176699 contig of 1581 bp in length
178681 contig of 1882 bp in length
180124 contig of 1343 bp in length
182291 contig of 2067 bp in length
183683 contig of 1292 bp in length
185242 contig of 1459 bp in length
186827 contig of 1485 bp in length
188237 contig of 1310 bp in length

Sequence updated (06-Jan-2000)
Sequence updated (26-May-2000)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 14984: contig of 14984 bp in length
* 14985 15084: gap of 100 bp
* 15085 25511: contig of 10427 bp in length
* 25512 25611: gap of 100 bp
* 25612 35208: contig of 9597 bp in length
* 35209 35308: gap of 100 bp
* 35309 45403: contig of 10095 bp in length
* 45404 45503: gap of 100 bp
* 45504 54399: contig of 8896 bp in length
* 54400 54499: gap of 100 bp
* 54500 63206: contig of 8707 bp in length
* 63207 63306: gap of 100 bp
* 63307 70841: contig of 7535 bp in length
* 70842 70941: gap of 100 bp
* 70942 79569: contig of 9028 bp in length
* 79970 80069: gap of 100 bp
* 80070 88468: contig of 8399 bp in length
* 88469 88568: gap of 100 bp
* 88569 94989: contig of 6421 bp in length
* 94990 95089: gap of 100 bp
* 95090 100403: contig of 5314 bp in length
* 100404 100503: gap of 100 bp

```

100504 105342: contig of 4839 bp in length  
105343 105442: gap of 100 bp  
105443 110335: contig of 4893 bp in length  
110336 110435: gap of 100 bp  
110436 115335: contig of 4900 bp in length  
115336 115435: gap of 100 bp  
115436 119336: contig of 4101 bp in length  
119337 119636: gap of 100 bp  
119637 124058: contig of 4422 bp in length  
124059 124158: gap of 100 bp  
124159 128684: contig of 4526 bp in length  
128685 128784: gap of 100 bp  
128785 132388: contig of 3604 bp in length  
132389 132488: gap of 100 bp  
132489 136139: contig of 3651 bp in length  
136140 136239: gap of 100 bp  
136240 139193: contig of 2954 bp in length  
139194 139293: gap of 100 bp  
139294 142207: contig of 2914 bp in length  
142208 142307: gap of 100 bp  
142308 146335: contig of 4028 bp in length  
146336 146435: gap of 100 bp  
146436 149853: contig of 3418 bp in length  
149854 149953: gap of 100 bp  
149954 152741: contig of 2788 bp in length  
152742 152841: gap of 100 bp  
152842 154497: contig of 1656 bp in length  
154498 154597: gap of 100 bp  
154598 156939: contig of 2342 bp in length  
156940 157039: gap of 100 bp  
157040 157559: contig of 520 bp in length  
157560 157659: gap of 100 bp  
157660 159511: contig of 1852 bp in length  
159512 159611: gap of 100 bp  
159612 162263: contig of 2652 bp in length  
162264 162363: gap of 100 bp  
162364 164148: contig of 1785 bp in length  
164149 164248: gap of 100 bp  
164249 166170: contig of 1922 bp in length  
166171 166270: gap of 100 bp  
166271 168664: contig of 2394 bp in length  
168665 168764: gap of 100 bp  
168765 170677: contig of 1913 bp in length  
170678 170777: gap of 100 bp  
170778 173092: contig of 2315 bp in length  
173093 173192: gap of 100 bp  
173193 175018: contig of 1826 bp in length  
175019 175118: gap of 100 bp  
175119 176699: contig of 1581 bp in length  
176700 176799: gap of 100 bp  
176800 178681: contig of 1882 bp in length  
178682 178781: gap of 100 bp  
178782 180124: contig of 1343 bp in length  
180125 180224: gap of 100 bp  
180225 182291: contig of 2067 bp in length  
182292 182391: gap of 100 bp  
182392 183683: contig of 1292 bp in length  
183684 183783: gap of 100 bp  
183784 185242: contig of 1459 bp in length  
185243 185342: gap of 100 bp  
185343 186827: contig of 1485 bp in length  
186828 186927: gap of 100 bp  
186928 188237: contig of 1310 bp in length.

FEATURES  
Source  
1. .188237  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11q22"  
/clone="RP11-70717"  
1. .14984  
misc\_feature  
/note="assembly\_fragment"  
15085. .25511

/note="assembly\_fragment"

Query Match 70.3%; Score 23.2; DB 2; Length 188237;  
Best Local Similarity 89.3%; Pred. No. 6.9;  
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 5 gatccttattggtttacactgactg 32  
DB 109550 GATCCTTATTGTTTCTCTGACTG 109577

RESULT 5

AC078936 196695 bp DNA linear HTG 11-AUG-2000  
Homo sapiens chromosome 2 clone RP11-69G4 map 2, WORKING DRAFT  
SEQUENCE, 13 unordered pieces.  
AC078936  
AC078936.1 GI:9795655  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
human.  
ORGANISM  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 196695)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 2, clone RP11-69G4  
Unpublished  
2 (bases 1 to 196695)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,  
Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Callymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehotzky,J.,  
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Melgrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenda,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (11-AUG-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Center project name: L9969  
Center clone name: 69.G.4  
----- Summary Statistics -----  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 191343 bases at least Q40  
Consensus quality: 194310 bases at least Q30  
Consensus quality: 195161 bases at least Q20  
Insert size: 185000; agarose-fp  
Insert size: 195495; sum-of-ctnigs  
Quality coverage: 5.4 in Q20 bases; agarose-fp

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Center project name: L9969  
Center clone name: 69.G.4  
----- Summary Statistics -----  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 191343 bases at least Q40  
Consensus quality: 194310 bases at least Q30  
Consensus quality: 195161 bases at least Q20  
Insert size: 185000; agarose-fp  
Insert size: 195495; sum-of-ctnigs  
Quality coverage: 5.4 in Q20 bases; agarose-fp

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Center project name: L9969  
Center clone name: 69.G.4  
----- Summary Statistics -----  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 191343 bases at least Q40  
Consensus quality: 194310 bases at least Q30  
Consensus quality: 195161 bases at least Q20  
Insert size: 185000; agarose-fp  
Insert size: 195495; sum-of-ctnigs  
Quality coverage: 5.4 in Q20 bases; agarose-fp

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Center project name: L9969  
Center clone name: 69.G.4  
----- Summary Statistics -----  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 191343 bases at least Q40  
Consensus quality: 194310 bases at least Q30  
Consensus quality: 195161 bases at least Q20  
Insert size: 185000; agarose-fp  
Insert size: 195495; sum-of-ctnigs  
Quality coverage: 5.4 in Q20 bases; agarose-fp

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Center project name: L9969  
Center clone name: 69.G.4  
----- Summary Statistics -----  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 191343 bases at least Q40  
Consensus quality: 194310 bases at least Q30  
Consensus quality: 195161 bases at least Q20  
Insert size: 185000; agarose-fp  
Insert size: 195495; sum-of-ctnigs  
Quality coverage: 5.4 in Q20 bases; agarose-fp



```

100722 113809 contig of 13088 bp in length
113910 117944 contig of 4035 bp in length
118045 121621 contig of 3577 bp in length
121722 126441 contig of 4720 bp in length
126542 130528 contig of 3987 bp in length
130629 138556 contig of 8228 bp in length
138957 146999 contig of 8043 bp in length
147000 150987 contig of 100 bp
148143 159863 contig of 9864 bp in length
150988 156963 contig of 100 bp
157064 165962 contig of 8899 bp in length
166063 172958 contig of 100 bp
173059 179585 contig of 6896 bp in length
175086 181934 contig of 1927 bp in length
182035 188042 contig of 6008 bp in length
188143 190987 contig of 100 bp
190988 196116 contig of 100 bp
191088 196216 contig of 100 bp
196217 200693 contig of 4477 bp in length
200694 200793 contig of 100 bp
200794 204728 contig of 3935 bp in length
204729 204828 contig of 100 bp
204829 208045 contig of 3217 bp in length
208046 208145 contig of 100 bp
208146 211951 contig of 3806 bp in length
211952 212051 contig of 100 bp
212052 215963 contig of 3912 bp in length
215964 216063 contig of 100 bp
216064 217202 contig of 1139 bp in length
217203 217302 contig of 100 bp
217303 220289 contig of 2987 bp in length
220290 220389 contig of 100 bp
220390 222754 contig of 2365 bp in length
222755 222854 contig of 100 bp
222855 225838 contig of 2984 bp in length
225839 225939 contig of 2518 bp in length
225939 228456 contig of 100 bp
228457 228556 contig of 2518 bp in length
228557 231507 contig of 2951 bp in length
231508 231607 contig of 100 bp
231608 234284 contig of 2677 bp in length
234285 236143 contig of 100 bp
236144 236243 contig of 1759 bp in length
236244 238375 contig of 100 bp
238376 238475 contig of 2132 bp in length
238476 240676 contig of 100 bp
240677 240776 contig of 2201 bp in length
240777 242833 contig of 100 bp
242834 242933 contig of 2057 bp in length
242934 244625 contig of 100 bp
244626 244725 contig of 1692 bp in length
244726 245870 contig of 100 bp
245871 245970 contig of 1145 bp in length
245971 248069 contig of 100 bp
248070 248169 contig of 2099 bp in length
248170 250179 contig of 100 bp
250180 250279 contig of 2010 bp in length
250280 250379 contig of 100 bp

```

Query Match 70.3%; Score 23.2; DB 2; Length 269711;  
 Best Local Similarity 89.3%; Pred. No. 6.8;  
 Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 56 contigs, the true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 19653: contig of 19653 bp in length
19654 19753: gap of 100 bp
19754 39362: contig of 19609 bp in length
39363 39462: gap of 100 bp
39463 55717: contig of 16255 bp in length
55718 55817: gap of 100 bp
55818 67507: contig of 11690 bp in length
67508 67607: gap of 100 bp
67608 77531: contig of 9924 bp in length
77532 77631: gap of 100 bp
77632 87218: contig of 9587 bp in length
87219 87318: gap of 100 bp
87319 95944: contig of 8626 bp in length
95945 96044: gap of 100 bp
96045 100521: contig of 4577 bp in length
100522 100721: gap of 100 bp
100722 113809: contig of 13088 bp in length

```

```

RESULT      8
SYNHYMED   SYNHYMED
LOCUS       SYNHYMED linear          SYN 27-APR-1993
DEFINITION  Synthetic hirudin gene, 3' end.
ACCESSION   M14964
VERSION     M14964.1 GI:208480
KEYWORDS    hirudin.
SOURCE      Synthetic DNA.
ORGANISM    synthetic construct
            artificial sequence.
REFERENCE   1 (bases 1 to 220)
AUTHORS     Fortkamp,E., Rieger,M., Heisterberg-Moutses,G., Schweitzer,S. and
            Sommer,R.
TITLE       Cloning and expression in Escherichia coli of a synthetic DNA for
            hirudin, the blood coagulation inhibitor in the leech
JOURNAL     DNA 5, 511-517 (1986)
MEDLINE     87132924
COMMENT     Draft entry and printed copy of sequence for [1] kindly provided by
            M.Rieger, 23-NOV-1988.
FEATURES
SOURCE
            1. 220
                Location/Qualifiers
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                <!. .207
                /note="hirudin precursor"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA72556.1"
                /db_xref="GI:208481"
                /translations="NSMVVYTDCTESGQNLCLCESNVCGOGNKCILGSDGKNQCQVT
                GEGTPKFSQSHNDGDFEIPFEIYLQ"
            10. .204
                /product="hirudin"
mat_peptide 61 a 50 c 59 g 50 t
            BASE COUNT 61 a 50 c 59 g 50 t
            ORIGIN
            ECORI site.

```

```

QY      10  ttatggtgtttacactgactgc 33
      | | | | | | | | | | | | | | |
Db      4  TCTATGGTTGTTTACACTGACTGC 27

      RESULT 10
      A00631/c
LOCUS      A00631          224 bp      DNA      linear      PAT 30-MAR-1993
DEFINITION Artificial sequence for hirudin-like protein, reverse complement.
ACCESSION  A00631
VERSION    A00631.1  GI:14562
KEYWORDS   .
SOURCE     synthetic construct.
ORGANISM   synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 224)
AUTHORS    .
TITLE      DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR
            PREPARING SUCH PROTEIN
JOURNAL    Patent: WO 8603517-A 16 19-JUN-1986;
FEATURES   Location/Qualifiers
            1..224
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT 51 a      60 c      51 g      62 t
ORIGIN

```

Query Match	67.9%	Score 22.4;	DB 6;	Length 224;
Best Local Similarity	95.8%	Pred. No. 24;		
Matches	23;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	10	tttatgggtgtttacactgactgc	33	
Db	221	TCTATGGTGTGTACACTGACTGC	198	
RESULT 11				
LOCUS	A04277	224 bp	DNA	linear
DEFINITION	Artificial sequence for hirudin-like fusion protein.			
ACCESSION	A04277			
VERSION	A04277.1	GI:344867		
KEYWORDS	synthetic construct.			
SOURCE	synthetic construct			
ORGANISM	artificial sequence.			
REFERENCE	1 (bases 1 to 224)			
AUTHORS	DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR			
TITLE	PREPARING SUCH PROTEIN			
JOURNAL	Patent: WO 8603517-A 19 JUN-1986;			
FEATURES	Location/Qualifiers			
source	1..224			
BASE COUNT	62 a	51 c	60 g	51 t
ORIGIN	/organism="synthetic construct"			
	/db_xref="taxon:32630"			
Query Match	67.9%	Score 22.4;	DB 6;	Length 224;
Best Local Similarity	95.8%	Pred. No. 24;		
Matches	23;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	10	tttatgggtgtttacactgactgc	33	
Db	4	TCTATGGTGTGTACACTGACTGC	27	
RESULT 12				
LOCUS	A04278/c	224 bp	DNA	linear
DEFINITION	Artificial sequence for hirudin-like fusion protein, reverse complement.			
ACCESSION	A04278			
VERSION	A04278.1	GI:344868		
KEYWORDS	synthetic construct.			
SOURCE	synthetic construct			
ORGANISM	artificial sequence.			
REFERENCE	1 (bases 1 to 224)			
AUTHORS	DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR			
TITLE	PREPARING SUCH PROTEIN			
JOURNAL	Patent: WO 8603517-A 20 JUN-1986;			
FEATURES	Location/Qualifiers			
source	1..224			
BASE COUNT	51 a	60 c	51 g	62 t
ORIGIN	/organism="synthetic construct"			
	/db_xref="taxon:32630"			
Query Match	67.9%	Score 22.4;	DB 6;	Length 224;
Best Local Similarity	95.8%	Pred. No. 24;		
Matches	23;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	10	tttatgggtgtttacactgactgc	33	
Db	221	TCTATGGTGTGTACACTGACTGC	198	

AUTHORS . DNA SEQUENCE ENCODING A HIRUDIN-LIKE PROTEIN AND PROCESS FOR  
 TITLE PREPARING SUCH PROTEIN  
 JOURNAL Patent: WO 8603517-A 2 19-JUN-1986;  
 FEATURES Location/Qualifiers  
 source 1..34  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 11 a 8 c 8 g 7 t  
 ORIGIN

Query Match 66.7%; Score 22; DB 6; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tatggtgtttacactgactgc 33  
 |||||  
 Db 33 TATGTTGTTTACTGACTGC 12

Search completed: July 15, 2002, 22:29:14  
 Job time: 19000 sec



GenCore version 4.5  
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:49 ; Search time 758.37 seconds  
(without alignments)  
74.710 Million cell updates/sec

Title: US-10-053-641-5

Perfect score: 33

Sequence: 1 tcgggaccccttattggtgtttacactgactgc 33

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22.4	67.9	224	7	AA60746
C 2	21.4	64.8	11472	23	ABL24190
3	21	63.6	726	12	AAQ12379
C 4	21	63.6	4809	22	AA53924
C 5	21	63.6	4809	22	AAH93831
6	20.8	63.0	217	7	AA60355
7	20.8	63.0	217	8	AA70323
8	20.2	61.2	9968	22	AA136353
9	19.6	59.4	36	19	AAV61848
					DNA encoding hirud
					Drosophila melanog
					Hirudin peptide/pr
					Human prostate cDN
					Human prostate-spe
					Desulphatohirudin
					Sequence of the de
					Human musculoskele
					A. contortrix prot

C 10	19.4	58.8	572	22	ABA14188	Human nervous syst
C 11	19.4	58.8	2362	22	AAH15864	Human cDNA sequenc
C 12	19.4	58.8	3112	21	AA538592	Human PRO701 prote
C 13	19.4	58.8	3113	20	AA234209	Human PRO701 nucle
C 14	19.4	58.8	3113	21	AA78552	Human PRO701 (UNQ3
C 15	19.4	58.8	3537	22	ABA21445	Human nervous syst
C 16	19.4	58.8	3539	22	ABA21444	Human nervous syst
C 17	19.4	58.8	10898	22	AAK86165	Human immune/haema
C 18	19.2	58.2	26767	23	ABL08654	Drosophila melanog
C 19	19.2	58.2	26960	23	ABL08712	Drosophila melanog
C 20	19	57.6	1323	22	AAH73251	Human cervical can
C 21	18.8	57.0	841	21	AA52083	Pigment protein fr
C 22	18.8	57.0	1193	21	AA36135	Arabidopsis thalia
C 23	18.8	57.0	1194	21	AA46472	Arabidopsis thalia
C 24	18.6	56.4	525	23	AA58147	cDNA #823 encoding
C 25	18.6	56.4	1846	22	AAH15263	Human cDNA sequenc
C 26	18.6	56.4	2837	23	AA587012	DNA encoding novel
C 27	18.6	56.4	3127	22	AAH54412	S. epidermidis gen
C 28	18.6	56.4	3157	22	AAH54304	S. epidermidis gen
C 29	18.6	56.4	17995	22	AAK85152	Human immune/haema
C 30	18.6	56.4	151826	21	AAF22291	BAC containing rep
C 31	18.4	55.8	300	21	AAA00372	Human colon cancer
C 32	18.4	55.8	2006	21	AA42229	Arabidopsis thalia
C 33	18.4	55.8	4968	23	ABL13204	Drosophila melanog
C 34	18.2	55.2	557	20	AA700474	Human secreted pro
C 35	18.2	55.2	567	22	AA530022	Human lung antigen
C 36	18.2	55.2	567	22	AA530023	Human lung antigen
C 37	18.2	55.2	567	22	AA528715	Genomic sequence #
C 38	18.2	55.2	567	22	AA528716	Genomic sequence #
C 39	18.2	55.2	830	22	AAK56695	Human immune/haema
C 40	18.2	55.2	7263	23	ABL23862	Drosophila melanog
C 41	18.2	55.2	8143	22	AA540435	DNA encoding human
C 42	18.2	55.2	8143	22	AA540435	Human reproductive
C 43	18.2	55.2	32191	22	AA526704	Human genomic DNA
C 44	18.2	55.2	49561	22	AAK82012	Human immune/haema
C 45	18	54.5	36	18	AA745950	Hirudin gene linke

## ALIGNMENTS

RESULT 1  
AA60746  
ID AA60746 standard; DNA; 224 BP.  
XX  
AC AA60746;  
XX  
AC AA60746;  
XX  
DT 01-JAN-1980 (first entry)  
XX  
DE DNA encoding hirudin-like polypeptide.  
XX  
KW Hirudin; thrombin-antagonist; anticoagulant; ds.  
XX  
OS Synthetic.  
XX  
PN DE3445517-A.  
XX  
PD 19-JUN-1986.  
XX  
PF 13-DEC-1984; 84DE-3445517.  
XX  
PR 13-DEC-1984; 84DE-3445517.  
XX  
PA (GENB-) GEN-BIO-TEC GES GEN.  
XX  
PI Fortkamp E, Rieger M, Sommer R;  
XX  
DR WPI; 1986-162802/26.  
XX  
DR P-PSDB; AAP60827.  
XX  
PT New DNA sequence coding for new hirudin like polypeptide - useful  
XX as thrombin antagonist, e.g. for inhibiting blood coagulation

PS Disclosure; Fig. 1; 25pp; German.

XX The sequence encodes a protein with hirudin-like activity. It is a

CC thrombin-antagonist useful for inhibiting blood coagulation and for

CC treating inflammation and/or oedema. It can be prepared in large

CC amounts, unlike natural hirudin which is available only in tiny

CC amounts from leeches. This polypeptide is produced using

CC recombinant DNA techniques and is expressed in a bacterial

CC transformant e.g. Escherichia coli.

XX Sequence 224 BP; 61 A; 51 C; 60 G; 51 T; 1 other;

SQ

Query Match 67.9%; Score 22.4; DB 7; Length 224;

Best Local Similarity 95.8%; Pred. No. 3.6;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 ttatggtgtttacactgactgc 33

DB 4 tctatggtgtttacactgactgc 27

RESULT 2

ABL24190/c

ID ABL24190 standard; DNA; 11472 BP.

XX

AC ABL24190;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 24043.

DE

DE Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

PI

XX WPI; 2001-656860/75.

DR

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Claim 1; SEQ ID NO 24043; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 11472 BP; 3419 A; 2447 C; 2370 G; 3236 T; 0 other;

Query Match 64.8%; Score 21.4; DB 23; Length 11472;

Best Local Similarity 80.6%; Pred. No. 16;

Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 tggggatccttatggtgtttacactgactgc 31

DB 10011 TCGGAGCGGTTATGTTGTTTCTTACT 9981

RESULT 3

AAQ12379

ID AAQ12379 standard; DNA; 726 BP.

XX

AC AAQ12379;

XX

DT 17-SEP-1991 (first entry)

XX

DE Hirudin peptide/Protein A fusion gene.

DE

KW anticoagulant; fusion protein; ds.

XX

XX Key Location/Qualifiers

FT mat\_peptide 1..489

FT /tag= a

FT /product= Protein A

FT 529..720

FT /tag= b

FT /product= Hirudin

FT 490..528

FT misc\_RNA

FT /tag= c

FT /product= linker oligopeptide Y

XX

XX DE3942580-A.

PN

XX 27-JUN-1991.

PD

XX 22-DEC-1989; 89DE-3942580.

PF

XX 22-DEC-1989; 89DE-3942580.

PR

XX (BADI ) BASF AG.

XX

XX Korwer W;

PI

XX WPI; 1991-194236/27.

DR

XX P-PSDB; AAR12751.

DR

XX Hirudin peptide prodn. by cleaving new fusion peptide - of

PT hirudin and protein A, expressed in high yield and stable,

PT soluble form by transformed E.coli

PT

XX Example 1; Page 6-7; 9pp; German.

PS

XX This sequence is an example of a fusion construct for expression of

CC the fusion peptide of the invention. The Met residue in the linker

CC oligopeptide allows cleavage by CNBr to release two fragments which

CC can be easily separated by IgG affinity chromatography. (The protein

CC A component binds to IgG sepharose). Increased yields of hirudin are

CC obtained using this fusion construct.

XX

SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match 63.6%; Score 21; DB 12; Length 726;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 atggtgtttacactgactgc 33

DB 523 atggtgtttacactgactgc 543

RESULT 4

PS Disclosure; Fig. 1; 25pp; German.

XX The sequence encodes a protein with hirudin-like activity. It is a

CC thrombin-antagonist useful for inhibiting blood coagulation and for

CC treating inflammation and/or oedema. It can be prepared in large

CC amounts, unlike natural hirudin which is available only in tiny

CC amounts from leeches. This polypeptide is produced using

CC recombinant DNA techniques and is expressed in a bacterial

CC transformant e.g. Escherichia coli.

XX Sequence 224 BP; 61 A; 51 C; 60 G; 51 T; 1 other;

SQ

Query Match 67.9%; Score 22.4; DB 7; Length 224;

Best Local Similarity 95.8%; Pred. No. 3.6;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 ttatggtgtttacactgactgc 33

DB 4 tctatggtgtttacactgactgc 27

RESULT 2

ABL24190/c

ID ABL24190 standard; DNA; 11472 BP.

XX

AC ABL24190;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 24043.

DE

DE Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

PI

XX WPI; 2001-656860/75.

DR

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Claim 1; SEQ ID NO 24043; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 11472 BP; 3419 A; 2447 C; 2370 G; 3236 T; 0 other;



XX OS Synthetic.  
XX Key Location/Qualifiers  
XX CDS 10..207  
XX FT /\*tag= a  
XX FT /product= hirudin  
XX PN EP168342-A.  
XX PD 15-JAN-1986.  
XX PF 10-JUN-1985; 85EP-0810268.  
XX PR 14-JUN-1984; 84CH-0288284.  
XX PA (CIBA ) CIBA GEIGY AG.  
XX PI Liersch M, Rink H, Markl W, Grutter MG, Meyhack B;  
XX WPI; 1986-015589/03.  
XX P-PSDB; AAP60395.  
XX DNA sequences coding for hirudin and derivs. - and expression  
XX PT vectors, transformed cells, monoclonal antibodies and hybridomas,  
XX PT useful as thrombin inhibitor.  
XX PS Disclosure; Page 10; 123pp; German.  
XX CC The sequence comprises an EcoRI restriction enzyme site at the  
XX CC 5'-end and a BamHI restriction enzyme site at the 3'-end.  
XX CC The sequence may be introduced into a vector for the transformation  
XX CC of hosts, e.g. E. coli. The hirudin can thus be prepared on a  
XX CC large scale. The product and its derivs. are thrombin inhibitors,  
XX CC useful in anticoagulant therapy, esp. when injected at doses  
XX CC of 0.01-0.05 mg/kg.  
XX SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;  
  
Query Match 63.0%; Score 20.8; DB 8; Length 217;  
Best Local Similarity 91.7%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 10 ttatggtgtgtttacactgactgc 33  
Db 7 ttcatggtgtgtttacacgactgc 30  
  
RESULT 8  
AAL36353  
ID AAL36353 standard; DNA; 9968 BP.  
XX AC AAL36353;  
XX DT 08-JAN-2002 (first entry)  
XX DE Human musculoskeletal system related polynucleotide SEQ ID NO 2718.  
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;  
XX KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX KW neurological disease; infection; human; secreted protein;  
XX KW musculoskeletal system; ds.  
XX OS Homo sapiens.  
XX PN WO200155367-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01338.  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 24-FEB-2000; 2000US-0184664.  
XX PR 02-MAR-2000; 2000US-0186350.  
XX PR 16-MAR-2000; 2000US-0189874.  
XX PR 17-MAR-2000; 2000US-0190076.  
XX PR 18-APR-2000; 2000US-0198123.  
XX PR 19-MAY-2000; 2000US-0205515.  
XX PR 07-JUN-2000; 2000US-0209467.  
XX PR 28-JUN-2000; 2000US-0214886.  
XX PR 30-JUN-2000; 2000US-0215135.  
XX PR 07-JUL-2000; 2000US-0216647.  
XX PR 07-JUL-2000; 2000US-0216880.  
XX PR 11-JUL-2000; 2000US-0217487.  
XX PR 14-JUL-2000; 2000US-0217496.  
XX PR 14-JUL-2000; 2000US-0218290.  
XX PR 26-JUL-2000; 2000US-0220963.  
XX PR 26-JUL-2000; 2000US-0220964.

XX OS Synthetic.  
XX Key Location/Qualifiers  
XX CDS 10..207  
XX FT /\*tag= a  
XX FT /product= hirudin  
XX PN EP168342-A.  
XX PD 15-JAN-1986.  
XX PF 10-JUN-1985; 85EP-0810268.  
XX PR 14-JUN-1984; 84CH-0288284.  
XX PA (CIBA ) CIBA GEIGY AG.  
XX PI Liersch M, Rink H, Markl W, Grutter MG, Meyhack B;  
XX WPI; 1986-015589/03.  
XX P-PSDB; AAP60395.  
XX DNA sequences coding for hirudin and derivs. - and expression  
XX PT vectors, transformed cells, monoclonal antibodies and hybridomas,  
XX PT useful as thrombin inhibitor.  
XX PS Disclosure; Page 10; 123pp; German.  
XX CC The sequence comprises an EcoRI restriction enzyme site at the  
XX CC 5'-end and a BamHI restriction enzyme site at the 3'-end.  
XX CC The sequence may be introduced into a vector for the transformation  
XX CC of hosts, e.g. E. coli. The hirudin can thus be prepared on a  
XX CC large scale. The product and its derivs. are thrombin inhibitors,  
XX CC useful in anticoagulant therapy, esp. when injected at doses  
XX CC of 0.01-0.05 mg/kg.  
XX SQ Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;  
  
Query Match 63.0%; Score 20.8; DB 7; Length 217;  
Best Local Similarity 91.7%; Pred. No. 17;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 10 ttatggtgtgtttacactgactgc 33  
Db 7 ttcatggtgtgtttacacgactgc 30  
  
RESULT 7  
AAN70323  
ID AAN70323 standard; DNA; 217 BP.  
XX AC AAN70323;  
XX DT 02-APR-1991 (first entry)  
XX DE Sequence of the desulphatohirudin gene.  
XX KW Anticoagulant; thrombin inhibitor; ds.  
XX PN EP225633-A.  
XX PD 16-JUN-1987.  
XX PF 09-DEC-1986; 86EP-0117098.  
XX PR 29-MAY-1986; 86GB-0013088.  
XX PR 12-DEC-1985; 85GB-0030631.  
XX PA (CIBA ) CIBA GEIGY AG.  
XX PA (PLAN-) PLANTORGAN WERK HEINRICH.  
XX PA (CHRI-) PLANTORGANW CHRISTENSEN.



CC and (f) infectious diseases such as viral, bacterial, fungal and  
 CC parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 9968 BP; 2324 A; 2302 C; 2695 G; 2647 T; 0 other;

Query Match 61.2%; Score 20.2; DB 22; Length 9968;  
 Best Local Similarity 88.0%; Pred. No. 52;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ggatccttattggtgttactg 28

Db 9864 ggatcgtttattggtttactg 9888

# RESULT 9

AAV61848  
 ID AAV61848 standard; DNA; 36 BP.

AC AAV61848;

XX 11-DEC-1998 (first entry)

DE A. contortrix protein C activator PCR primer #3.

KW Protein C activator; anticoagulant; membrane-bound; inhibitor; clotting;  
 KW endothelial cell; organ rejection; transplantation; xenotransplantation.  
 KW PCR primer; ss.

XX Synthetic.

OS Agkistrodon contortrix.

XX WO9842850-A1.

XX 01-OCT-1998.

XX 26-MAR-1998; 98WO-GB008050.

XX 23-SEP-1997; 97GB-0020248.

PR 26-MAR-1997; 97GB-0006327.

XX (RPMS-) RPMS TECHNOLOGY LTD.

XX Dorling A, George AJT, Lechler RI, Riesbeck K;

XX WPI; 1998-557034/47.

PT New fusion proteins used particularly for preventing transplant  
 PT rejection - comprising a region with anticoagulant activity and a  
 PT region which can anchor the protein to a cell membrane

XX Disclosure; Page 29; 67pp; English.

CC AAV61846-V61860 are PCR primers used to amplify a protein C activator  
 CC from Agkistrodon contortrix. This protein contains a region with  
 CC anticoagulant activity and a region which can anchor the protein to a  
 CC cell membrane. The protein can provide membrane-bound anticoagulant  
 CC proteins. They are suitable for inhibiting the clotting cascade at the  
 CC surface of endothelial cells (ECs) thus inhibiting in vivo mechanisms  
 CC responsible for organ rejection. They can provide biological tissue  
 CC suitable for transplantation, particularly for xenotransplantation.

XX Sequence 36 BP; 7 A; 12 C; 9 G; 8 T; 0 other;

# Query Match

Best Local Similarity 59.4%; Score 19.6; DB 19; Length 36;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 cctttatggtgtttactgactgac 33

Db 8 cctggcggtgtttactgactgac 33

# RESULT 10

ABAL4188/C

ID ABAL4188 standard; cDNA; 572 BP.

XX ABAL4188;

XX 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 3195.

KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antischistosomal; antianaemic; antithrombotic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.

XX Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180828.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0215135.

PR 30-JUN-2000; 2000US-0218866.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 18-AUG-2000; 2000US-0225759.

PR 22-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-02411785.  
PR 20-OCT-2000; 2000US-02411786.  
PR 20-OCT-2000; 2000US-02411787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2000US-0254097.  
PR 05-JAN-2001; 2000US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPT; 2001-541565/60.  
DR P-FSDB; ABB17862.  
XX  
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
PS Claim 1; SEQ ID NO 3195; 1701pp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
CC (ABAI14678-ABAI18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 572 BP; 171 A; 169 C; 106 G; 119 T; 7 other;

Query Match 58.8%; Score 19.4; DB 22; Length 572;  
Best Local Similarity 79.3%; Pred. No. 78;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 4 ggatccttattggtgtttacactgactg 32  
||| |||| ||||| ||| ||||  
Db 223 GGAACCTTTGTGTTGTTGAACACTACTG 195

RESULT 11  
AAH15864/c  
ID AAH15864 standard; cDNA; 2362 BP.  
XX  
AC AAH15864;  
XX  
DT 26-JUN-2001 (first entry)  
XX

DE Human cDNA sequence SEQ ID NO:14380.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
KW Homo sapiens.  
OS EP1074617-A2.  
XX  
PN 07-FEB-2001.  
XX  
PD 28-JUL-2000; 2000EP-0116126.  
XX  
PF 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 14380; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX Sequence 2362 BP; 654 A; 671 C; 534 G; 503 T; 0 other;  
SQ

Query Match 58.8%; Score 19.4; DB 22; Length 2362;  
Best Local Similarity 79.3%; Pred. No. 94;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 4 ggatccttatgggtgtttacactgactg 32  
||| ||||| ||||| ||||| ||||| |||||  
Db 1139 GGAACCTTTGTGTTGTTGAACATACATG 1111

RESULT 12  
AAC58592/C  
ID AAC58592 standard; cDNA; 3112 BP.  
XX  
AC AAC58592;  
XX

DT 29-JAN-2001 (first entry)  
XX Human PRO701 protein UNQ365 encoding cDNA SEQ ID NO:66.  
DE  
XX Human; immune related disease; diagnosis; antiinflammatory; cardiant;  
KW dermatological; antiarthritic; antirheumatic; immunosuppressive;  
KW haemostatic; antithyroid; antidiabetic; nootropic; neuroprotective;  
KW antianaemic; hepatotropic; virucide; antipsoriatic; antiallergic;  
KW osteoarthritis; systemic lupus erythematosus; rheumatoid arthritis;  
KW idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;  
KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;  
KW autoimmune thrombocytopenia; immune-mediated renal disease;  
KW demyelinating disease; hepatobiliary disease; Whipple's disease;  
KW inflammatory bowel disease; gluten-sensitive enteropathy;  
KW autoimmune disease; immune-mediated skin disease; allergic disease;  
KW immunological disease; transplantation associated disease;  
KW graft rejection; graft-versus-host-disease; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200053758-A2.  
PN  
XX 14-SEP-2000.  
PD  
XX  
XX 02-MAR-2000; 2000WO-US05841.  
PF  
XX 08-MAR-1999; 99WO-US05028.  
PR 10-MAR-1999; 99US-0123618.  
PR 12-MAR-1999; 99US-0123957.  
PR 23-MAR-1999; 99US-0125775.  
PR 12-APR-1999; 99US-0128849.  
PR 20-APR-1999; 99WO-US08615.  
PR 28-APR-1999; 99US-0131445.  
PR 04-MAY-1999; 99US-0132371.  
PR 14-MAY-1999; 99US-0134287.  
PR 02-JUN-1999; 99WO-US12252.  
PR 23-JUN-1999; 99US-0141037.  
PR 20-JUL-1999; 99US-0144758.  
PR 26-JUL-1999; 99US-0145698.  
PR 28-JUL-1999; 99US-0146222.  
PR 01-SEP-1999; 99WO-US20111.  
PR 08-SEP-1999; 99WO-US20594.  
PR 13-SEP-1999; 99WO-US20944.  
PR 15-SEP-1999; 99WO-US21090.  
PR 15-SEP-1999; 99WO-US21547.  
PR 05-OCT-1999; 99WO-US23089.  
PR 29-OCT-1999; 99US-0162506.  
PR 30-NOV-1999; 99WO-US28214.  
PR 30-NOV-1999; 99WO-US28313.  
PR 30-NOV-1999; 99WO-US28409.  
PR 01-DEC-1999; 99WO-US28301.  
PR 01-DEC-1999; 99WO-US28634.  
PR 02-DEC-1999; 99WO-US28551.  
PR 02-DEC-1999; 99WO-US28564.  
PR 16-DEC-1999; 99WO-US28565.  
PR 16-DEC-1999; 99WO-US30095.  
PR 20-DEC-1999; 99WO-US30999.  
PR 30-DEC-1999; 99WO-US31274.  
PR 05-JAN-2000; 2000WO-US00219.  
PR 06-JAN-2000; 2000WO-US00277.  
PR 06-JAN-2000; 2000WO-US00376.  
PR 11-FEB-2000; 2000WO-US03565.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 18-FEB-2000; 2000WO-US04342.  
PR 22-FEB-2000; 2000WO-US04414.  
XX  
XX (GETH ) GENENTECH INC.  
XX Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W;  
PI Kabakoff RC, Lu Y, Pan J, Pennica D, Shelton DL, Smith V;  
PI Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M;



DR WPI: 2000-572271/53.  
 DR P-PSDB; AAB33427.  
 XX  
 PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of  
 PT immune related disorders, e.g. systemic lupus erythematosus, rheumatoid  
 PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -  
 XX  
 XX Claim 23; Fig 27; 309pp; English.  
 XX  
 CC The present invention describes sixty four human PRO proteins which can  
 CC be used in the treatment of immune related diseases. The human PRO  
 CC proteins, anti-PRO antibodies, agonists and antagonists are useful for  
 CC treating and diagnosing immune related disorders. The disorders are  
 CC selected from systemic lupus erythematosus, rheumatoid arthritis,  
 CC osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,  
 CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's  
 CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic  
 CC anaemia, autoimmune thrombocytopaenia, thyroiditis, diabetes mellitus,  
 CC and peripheral nervous systems, demyelinating diseases of the central  
 CC bowel disease, gluten-sensitive enteropathy and whipple's disease,  
 CC autoimmune or immune-mediated skin diseases, allergic diseases,  
 CC immunological diseases of the lung, and transplantation associated  
 CC diseases including graft rejection and graft-versus-host-disease.  
 CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used  
 CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and  
 CC AAB33414 to AAB33477 represent human PRO polynucleotide and protein  
 CC sequences given in the exemplification of the present invention.  
 XX  
 SQ Sequence 3112 BP; 853 A; 867 C; 712 G; 580 T; 0 other;

Query Match 58.8%; Score 19.4; DB 21; Length 3112;  
 Best Local Similarity 79.3%; Pred. No. 97;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 ggatcctttatgtgtgttttaactgactg 32  
 ||| ||||| ||||| ||| |||||

DB 1890 GGAACTTTGTTGGTTGTAACATACG 1862

RESULT 13

AAZ34209/C  
 ID AAZ34209 standard; cDNA; 3113 BP.

XX AC AAZ34209;

XX DT 07-DEC-1999 (first entry)

XX DE Human PRO701 nucleotide sequence.

XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
 KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
 KW secreted protein; transmembrane protein; ss.

OS Homo sapiens.

XX PN W09946281-A2.

XX PD 16-SEP-1999.

XX PF 08-MAR-1999; 99WO-US05028.

XX 10-MAR-1998; 98US-0077450.

XX 11-MAR-1998; 98US-0077632.

XX 11-MAR-1998; 98US-0077641.

XX 11-MAR-1998; 98US-0077649.

XX 12-MAR-1998; 98US-0077791.

XX 13-MAR-1998; 98US-0078004.

XX 17-MAR-1998; 98US-0040220.

XX 20-MAR-1998; 98US-0078886.

XX 20-MAR-1998; 98US-0078910.

XX 20-MAR-1998; 98US-0078936.

PR 20-MAR-1998; 98US-0078939.  
 PR 25-MAR-1998; 98US-0079294.  
 PR 26-MAR-1998; 98US-0079656.  
 PR 27-MAR-1998; 98US-0079663.  
 PR 27-MAR-1998; 98US-0079664.  
 PR 27-MAR-1998; 98US-0079689.  
 PR 27-MAR-1998; 98US-0079728.  
 PR 27-MAR-1998; 98US-0079786.  
 PR 30-MAR-1998; 98US-0079920.  
 PR 30-MAR-1998; 98US-0079923.  
 PR 31-MAR-1998; 98US-0080105.  
 PR 31-MAR-1998; 98US-0080107.  
 PR 31-MAR-1998; 98US-0080165.  
 PR 31-MAR-1998; 98US-0080194.  
 PR 01-APR-1998; 98US-0080327.  
 PR 01-APR-1998; 98US-0080328.  
 PR 01-APR-1998; 98US-0080333.  
 PR 01-APR-1998; 98US-0080334.  
 PR 08-APR-1998; 98US-0081049.  
 PR 08-APR-1998; 98US-0081070.  
 PR 08-APR-1998; 98US-0081071.  
 PR 09-APR-1998; 98US-0081195.  
 PR 09-APR-1998; 98US-0081203.  
 PR 09-APR-1998; 98US-0081229.  
 PR 15-APR-1998; 98US-0081817.  
 PR 15-APR-1998; 98US-0081838.  
 PR 15-APR-1998; 98US-0081952.  
 PR 15-APR-1998; 98US-0081955.  
 PR 21-APR-1998; 98US-0082568.  
 PR 21-APR-1998; 98US-0082569.  
 PR 22-APR-1998; 98US-0082700.  
 PR 22-APR-1998; 98US-0082704.  
 PR 23-APR-1998; 98US-0082804.  
 PR 23-APR-1998; 98US-0082767.  
 PR 23-APR-1998; 98US-0082796.  
 PR 27-APR-1998; 98US-0083336.  
 PR 28-APR-1998; 98US-0083322.  
 PR 29-APR-1998; 98US-0083392.  
 PR 29-APR-1998; 98US-0083495.  
 PR 29-APR-1998; 98US-0083496.  
 PR 29-APR-1998; 98US-0083499.  
 PR 29-APR-1998; 98US-0083500.  
 PR 29-APR-1998; 98US-0083545.  
 PR 29-APR-1998; 98US-0083554.  
 PR 29-APR-1998; 98US-0083558.  
 PR 29-APR-1998; 98US-0083559.  
 PR 30-APR-1998; 98US-0083742.  
 PR 03-MAY-1998; 98US-0084366.  
 PR 06-MAY-1998; 98US-0084414.  
 PR 06-MAY-1998; 98US-0084441.  
 PR 07-MAY-1998; 98US-0084598.  
 PR 07-MAY-1998; 98US-0084600.  
 PR 07-MAY-1998; 98US-0084627.  
 PR 07-MAY-1998; 98US-0084637.  
 PR 07-MAY-1998; 98US-0084639.  
 PR 07-MAY-1998; 98US-0084640.  
 PR 07-MAY-1998; 98US-0084643.  
 PR 13-MAY-1998; 98US-0085323.  
 PR 13-MAY-1998; 98US-0085338.  
 PR 13-MAY-1998; 98US-0085339.  
 PR 15-MAY-1998; 98US-0085573.  
 PR 15-MAY-1998; 98US-0085579.  
 PR 15-MAY-1998; 98US-0085580.  
 PR 15-MAY-1998; 98US-0085582.  
 PR 15-MAY-1998; 98US-0085689.  
 PR 15-MAY-1998; 98US-0085697.  
 PR 15-MAY-1998; 98US-0085700.  
 PR 18-MAY-1998; 98US-0085704.  
 PR 22-MAY-1998; 98US-0086023.  
 PR 22-MAY-1998; 98US-0086392.  
 PR 22-MAY-1998; 98US-0086414.  
 PR 22-MAY-1998; 98US-0086430.  
 PR 22-MAY-1998; 98US-0086486.

PR 28-MAY-1998; 98US-0087098.  
 PR 28-MAY-1998; 98US-0087106.  
 PR 28-MAY-1998; 98US-0087208.  
 PR 30-JUL-1998; 98US-0094651.  
 PR 11-SEP-1998; 98US-0100038.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX Wood WL, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;  
 XX WPI; 1999-551358/46.  
 DR P-PSDB; AAY41740.  
 XX  
 PT New secreted and transmembrane polypeptides and their polynucleotides,  
 PT useful for treating blood coagulation disorders, cancers and cellular  
 PT adhesion disorders -  
 XX  
 XX Claim 2; Fig 150; 530pp; English.  
 PS  
 XX The present invention describes secreted and transmembrane polypeptides  
 CC and their polynucleotides. The nucleotide sequences are useful as  
 CC sources of probes, primers, for chromosome mapping, and for generation  
 CC of antisense sequences. They can also be used to create transgenic  
 CC animals. The proteins can be used to treat a variety of diseases and  
 CC disorders, depending on their function. Diseases that may be treated  
 CC include blood coagulation disorders, cancers and cellular adhesion  
 CC disorders. They may also be used to raise antibodies. AAZ33891 to  
 CC AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and  
 CC polypeptide sequence given in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 3113 BP; 853 A; 867 C; 712 G; 681 T; 0 other;

Query Match 58.8%; Score 19.4; DB 20; Length 3113;  
 Best Local Similarity 79.3%; Pred. No. 97;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 ggatccttattggtgtttacactgactg 32  
 ||| ||||| ||||| || |||||  
 Db 1890 GGAACCTTTGTTGTTGTTGAACATACCTG 1862

RESULT 14  
 AAC78552/G  
 ID AAC78552 standard; cDNA; 3113 BP.  
 XX  
 AC AAC78552;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human PRO701 (UNQ365) nucleotide sequence SEQ ID NO:374.  
 XX  
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;  
 KW expressed sequence tag; detection; cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200053756-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 18-FEB-2000; 2000WO-US04341.  
 XX  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 29-MAR-1999; 99US-0126773.  
 PR 21-APR-1999; 99US-0130232.  
 PR 28-APR-1999; 99US-0131445.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 XX  
 XX (GETH ) GENENTECH INC.  
 PA Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
 XX Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;  
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
 PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;  
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WL;  
 XX  
 DR WPI; 2000-611443/58.  
 DR P-PSDB; AAB44296.  
 DR  
 XX Novel PRO polypeptides and polynucleotides used in detection methods,  
 PT to target bioactive molecules to specific cells, and to modulate  
 PT cellular activities -  
 XX  
 PS Claim 2; Fig 150; 636pp; English.  
 XX  
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed  
 CC sequence tag) sequences which encode secreted or transmembrane PRO  
 CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
 CC activity. The polynucleotides and polypeptides can be used for detecting  
 CC the presence of PRO polypeptides in samples, for linking bioactive  
 CC molecules to cells and for modulating biological activities of cells,  
 CC using the polypeptides for specific targeting. The polypeptide targeting  
 CC can be used to kill the target cells, e.g. for the treatment of cancers.  
 CC The polypeptide pairs provide specific targeting of bioactive molecules  
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in  
 CC the isolation of the PRO polynucleotide sequences.  
 XX  
 SQ Sequence 3113 BP; 853 A; 867 C; 712 G; 681 T; 0 other;

Query Match 58.8%; Score 19.4; DB 21; Length 3113;  
 Best Local Similarity 79.3%; Pred. No. 97;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 ggatccttattggtgtttacactgactg 32  
 ||| ||||| ||||| || |||||  
 Db 1890 GGAACCTTTGTTGTTGTTGAACATACCTG 1862

RESULT 15  
 ABA21445/c  
 ID ABA21445 standard; DNA; 3537 BP.  
 XX  
 AC ABA21445;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Human nervous system related polynucleotide SEQ ID NO 13776.  
 XX  
 KW Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antistickling; antianemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 OS Homo sapiens.  
 XX  
 PN WO200159063-A2.  
 XX

PD	16-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01334.		
XX			
XX	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	09-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226868.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR	06-SEP-2000; 2000US-0230437.		
PR	06-SEP-2000; 2000US-0230438.		
PR	08-SEP-2000; 2000US-0231242.		
PR	08-SEP-2000; 2000US-0231243.		
PR	08-SEP-2000; 2000US-0231244.		
PR	08-SEP-2000; 2000US-0231413.		
PR	08-SEP-2000; 2000US-0231414.		
PR	08-SEP-2000; 2000US-0232080.		
PR	08-SEP-2000; 2000US-0232081.		
PR	12-SEP-2000; 2000US-0231968.		
PR	14-SEP-2000; 2000US-0232397.		
PR	14-SEP-2000; 2000US-0232398.		
PR	14-SEP-2000; 2000US-0232399.		
PR	14-SEP-2000; 2000US-0232400.		
PR	14-SEP-2000; 2000US-0232401.		
PR	14-SEP-2000; 2000US-0233063.		
PR	14-SEP-2000; 2000US-0233064.		
PR	14-SEP-2000; 2000US-0233065.		
PR	21-SEP-2000; 2000US-0233066.		
PR	21-SEP-2000; 2000US-0234223.		
PR	25-SEP-2000; 2000US-0234997.		
PR	25-SEP-2000; 2000US-0234998.		
PR	26-SEP-2000; 2000US-0235484.		
PR	27-SEP-2000; 2000US-0235834.		
PR	27-SEP-2000; 2000US-0235836.		
PR	29-SEP-2000; 2000US-0236327.		
PR	29-SEP-2000; 2000US-0236367.		
PR	29-SEP-2000; 2000US-0236368.		
PR	29-SEP-2000; 2000US-0236369.		
PR	29-SEP-2000; 2000US-0236370.		
PR	02-OCT-2000; 2000US-0236802.		
PR	02-OCT-2000; 2000US-0237037.		
PR	02-OCT-2000; 2000US-0237038.		
PR	02-OCT-2000; 2000US-0237039.		
PR	02-OCT-2000; 2000US-0237040.		
PR	13-OCT-2000; 2000US-0239935.		
PR	13-OCT-2000; 2000US-0239937.		
PR	20-OCT-2000; 2000US-0240960.		
PR	20-OCT-2000; 2000US-0241785.		
PR	20-OCT-2000; 2000US-0241786.		
PR	20-OCT-2000; 2000US-0241787.		
PR	20-OCT-2000; 2000US-0241808.		
PR	20-OCT-2000; 2000US-0241809.		
PR	20-OCT-2000; 2000US-0241826.		
PR	20-OCT-2000; 2000US-0242221.		
PR	01-NOV-2000; 2000US-0244617.		
PR	08-NOV-2000; 2000US-0246474.		
PR	08-NOV-2000; 2000US-0246475.		
PR	08-NOV-2000; 2000US-0246476.		
PR	08-NOV-2000; 2000US-0246477.		
PR	08-NOV-2000; 2000US-0246478.		
PR	08-NOV-2000; 2000US-0246523.		
PR	08-NOV-2000; 2000US-0246524.		
PR	08-NOV-2000; 2000US-0246525.		
PR	08-NOV-2000; 2000US-0246526.		
PR	08-NOV-2000; 2000US-0246527.		
PR	08-NOV-2000; 2000US-0246528.		
PR	08-NOV-2000; 2000US-0246532.		
PR	08-NOV-2000; 2000US-0246609.		
PR	08-NOV-2000; 2000US-0246610.		
PR	08-NOV-2000; 2000US-0246611.		
PR	08-NOV-2000; 2000US-0246613.		
PR	17-NOV-2000; 2000US-0249207.		
PR	17-NOV-2000; 2000US-0249208.		
PR	17-NOV-2000; 2000US-0249209.		
PR	17-NOV-2000; 2000US-0249210.		
PR	17-NOV-2000; 2000US-0249211.		
PR	17-NOV-2000; 2000US-0249212.		
PR	17-NOV-2000; 2000US-0249213.		
PR	17-NOV-2000; 2000US-0249214.		
PR	17-NOV-2000; 2000US-0249215.		
PR	17-NOV-2000; 2000US-0249216.		
PR	17-NOV-2000; 2000US-0249217.		
PR	17-NOV-2000; 2000US-0249218.		
PR	17-NOV-2000; 2000US-0249244.		
PR	17-NOV-2000; 2000US-0249245.		
PR	17-NOV-2000; 2000US-0249264.		
PR	17-NOV-2000; 2000US-0249265.		
PR	17-NOV-2000; 2000US-0249267.		
PR	17-NOV-2000; 2000US-0249297.		
PR	17-NOV-2000; 2000US-0249299.		
PR	17-NOV-2000; 2000US-0249300.		
PR	01-DEC-2000; 2000US-0250391.		
PR	01-DEC-2000; 2000US-0251160.		
PR	05-DEC-2000; 2000US-0251030.		
PR	05-DEC-2000; 2000US-0251988.		
PR	05-DEC-2000; 2000US-0256719.		
PR	06-DEC-2000; 2000US-0251479.		
PR	08-DEC-2000; 2000US-0251856.		
PR	08-DEC-2000; 2000US-0251868.		
PR	08-DEC-2000; 2000US-0251869.		
PR	08-DEC-2000; 2000US-0251989.		
PR	08-DEC-2000; 2000US-0251990.		
PR	11-DEC-2000; 2000US-0254097.		
PR	05-JAN-2001; 2001US-0259678.		
XX			
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA			
XX	Rosen CA, Barash SC, Ruben SM;		
PI			
XX	WPI; 2001-541565/60.		
XX			
DR			

XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating nervous system  
 PT cancers and metastases -  
 XX  
 XX Disclosure; SEQ ID NO 13776; 1701pp + Sequence Listing; English.  
 PS  
 XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
 CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 3537 BP; 1142 A; 722 C; 649 G; 1024 T; 0 other;

Query Match 58.8%; Score 19.4; DB 22; Length 3537;  
 Best Local Similarity 79.3%; Pred. No. 99;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Qy 4 ggatcctttatggtgtttacactgactg 32  
 ||| ||||| ||||| || |||||  
 Db 223 GGAACCTTGTGTGTGTGAACATACTG 195

Search completed: July 15, 2002, 22:44:53  
 Job time: 10066 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:31:49 ; Search time 174.99 Seconds  
(without alignments)  
46.322 Million cell updates/sec

Title: US-10-053-641-5

Perfect score: 33

Sequence: 1 tcgggaccccttattggtttacactgactgc 33

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 393533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	66.7	93	1	US-07-859-453E-3
2	21	63.6	106	1	US-08-262-384A-6
3	21	63.6	129	1	US-08-262-384A-3
4	21	63.6	235	1	US-08-262-384A-7
5	21	63.6	235	1	US-08-262-384A-8
6	21	63.6	726	1	US-08-262-384A-9
7	21	63.6	726	1	US-08-262-384A-10
8	20.8	63.0	58	6	5422249-10
9	20.8	63.0	109	6	5422249-6
10	20.8	63.0	217	6	5422249-13
11	20.8	63.0	218	6	5422249-4
12	19.4	58.8	122	1	US-08-619-598-4
13	18.6	56.4	5676	2	US-08-663-998-3
14	18.6	56.4	5682	2	US-08-663-998-4
15	18.4	55.8	41	1	US-07-854-596B-50
16	18.4	55.8	43	1	US-08-044-506B-9
17	18.4	55.8	44	1	US-07-854-598B-51
18	18.4	55.8	36	1	US-08-466-265-12
19	18.4	54.5	38	1	US-08-466-265-11
20	18.4	54.5	38	1	US-08-605-053-4
21	18.4	54.5	105	1	US-07-842-089E-24
22	18.4	54.5	105	1	US-08-264-485-24
23	17.8	53.9	227	6	5164304-3
24	17.8	53.9	228	6	5179196-3
25	17.4	52.7	39	1	US-08-044-506B-8
26	17.4	52.7	1107	1	US-08-292-945-5
27	17.4	52.7	1107	1	US-08-252-073A-5

Sequence 5, Appl  
Sequence 28, Appl  
Sequence 1, Appl  
Sequence 13, Appl  
Sequence 13, Appl  
Sequence 13, Appl  
Sequence 1, Appl  
Patent No. 5422249  
Sequence 19, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 23, Appl

1107 5 PCT-US93-12074-5  
36 4 US-08-235-836C-28  
1794 3 US-08-920-270-1  
2203 1 US-08-076-091C-13  
2203 1 US-08-285-641-13  
5055 4 US-09-242-632A-13  
2772 3 US-08-936-135-1  
201 6 5422249-3  
736 4 US-09-268-364-19  
1131 1 US-07-959-946-2  
1131 1 US-08-333-577-2  
1131 5 PCT-US92-08634-2  
2473 2 US-08-540-804-3  
2473 2 US-08-218-265-3  
2473 3 US-08-521-872-3  
2473 4 US-08-590-399-3  
3147 2 US-07-952-853-23  
3147 2 US-08-914-848-23

## ALIGNMENTS

## RESULT 1

US-07-859-453E-3  
; Sequence 3, Application US/07859453E  
; Patent No. 5663141

GENERAL INFORMATION:

APPLICANT: Kurfuerst, Manfred  
APPLICANT: Ruebsamen, Klaus  
APPLICANT: Schmid, Bernhard

APPLICANT: Koerwer, Wolfgang  
APPLICANT: Schwenen, Juergen

APPLICANT: Hoeftken, Hans Wolfgang

TITLE OF INVENTION: Hirudin/polyalkylene glycol

TITLE OF INVENTION: conjugates

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kell & Weinkauff

STREET: 1101 Connecticut Avenue

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

COMPUTER: IBM compatible

OPERATING SYSTEM: MS-DOS version 7.0

SOFTWARE: WordPerfect version 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/859,453E

FILING DATE: 29-MAY-1992

CLASSIFICATION: 514

CLASSIFICATION: A 61 K 37/64

CLASSIFICATION: A 61 K 47/48

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP90/01998

FILING DATE: 22-NOV-1990

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

US-07-859-453E-3

Query Match 66.7%; Score 22; DB 1; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tatgggtgtttacactgactgc 33  
|||||

us-10-053-641-5.rn1

Tue Jul 16 08:24:11 2002

```

Db 15 TATGTTTCTTACACTGACTGC 36

RESULT 2
US-08-262-384A-6/c
; Sequence 6, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-262-384A-3

Query Match 63.6%; Score 21; DB 1; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 atggtgttttacactgactgc 33
|||||
Db 31 ATGGTTGTTTACACTGACTGC 51

RESULT 4
US-08-262-384A-7
; Sequence 7, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-262-384A-7

Query Match 63.6%; Score 21; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 atggtgttttacactgactgc 33
|||||
Db 81 ATGGTTGTTTACACTGACTGC 61

RESULT 3
US-08-262-384A-3
; Sequence 3, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0

```

Query Match 63.6%; Score 21; DB 1; Length 235;  
Best Local Similarity 100.0%; Pred. No. 0.96;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 atgggtgtttacactgactgc 33  
|||||  
Db 31 ATGGTGTGTTACTGACTGTC 51

RESULT 5

US-08-262-384A-8/c  
; Sequence 8, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994

CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990

APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992

INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 235 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

US-08-262-384A-8

Query Match 63.6%; Score 21; DB 1; Length 235;  
Best Local Similarity 100.0%; Pred. No. 0.96;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 atgggtgtttacactgactgc 33  
|||||  
Db 209 ATGGTGTGTTACTGACTGTC 189

RESULT 6

US-08-262-384A-9  
; Sequence 9, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington

STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994

CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990

APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992

INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 726 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

US-08-262-384A-9

Query Match 63.6%; Score 21; DB 1; Length 726;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 atgggtgtttacactgactgc 33  
|||||  
Db 523 ATGGTGTGTTACTGACTGTC 543

RESULT 7

US-08-262-384A-10/c  
; Sequence 10, Application US/08262384A  
; Patent No. 5624822

GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.

COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994

CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990

APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992

INFORMATION FOR SEQ ID NO: 10:





Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 10 ttatggtgtttacactgactgc 33  
|| |||||  
Db 7 ttcattggtgtttacaccgactgc 30

## RESULT 12

US-08-619-598-4/c  
; Sequence 4, Application US/08619598  
; Patent No. 5672487

## ; GENERAL INFORMATION:

; APPLICANT: SCHWEDEN, Juergen  
; APPLICANT: BOLLSCHWEILER, Claus  
; APPLICANT: PIONTEK, Michael  
; APPLICANT: WEYDEMANN, Ulrike

; APPLICANT: GELLISSEN, Gerd

; TITLE OF INVENTION: THE RECOMBINANT PRODUCTION

; TITLE OF INVENTION: OF PROTEINS IN YEAST

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kell & Weinlauf

; STREET: 1101 Connecticut Avenue

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

; COMPUTER: IBM AT-compatible, 80486 processor

; OPERATING SYSTEM: MS-DOS version 6.0

; SOFTWARE: WordPerfect version 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/619,598

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP/94/03409

; FILING DATE: 17-OCT-1994

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 122 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

US-08-619-598-4

Query Match 58.8%; Score 19.4; DB 1; Length 122;

Best Local Similarity 95.2%; Pred. No. 4.4;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 13 atggtgtttacactgactgc 33

|||||

Db 35 AGGCTGTTTACACTGACTGC 15

## RESULT 13

US-08-663-998-3/c

; Sequence 3, Application US/08663998

; Patent No. 5846946

## ; GENERAL INFORMATION:

; APPLICANT: HUEBNER, Robert C.

; APPLICANT: NORMAN, Jon A.

; APPLICANT: LIANG, Xiaowu

; APPLICANT: CARNER, Kristin R.

; APPLICANT: BARBOUR, Alan G.

; APPLICANT: LUKE, Catherine J.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/663,998

; FILING DATE: 06-JUN-1996

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Kowalski, Thomas J.

; REGISTRATION NUMBER: 32,147

; REFERENCE/DOCKET NUMBER: 454312-2440

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-840-3333

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5676 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-663-998-3

Query Match 56.4%; Score 18.6; DB 2; Length 5676;

Best Local Similarity 72.7%; Pred. No. 22;

Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 tcgggatccttattggtgtttacactgactgc 33

|||||

Db 2448 TCTGGATCCTTAAGGTTTTTTTGGACTTCTGC 2416

## RESULT 14

US-08-663-998-4/c

; Sequence 4, Application US/08663998

; Patent No. 5846946

## ; GENERAL INFORMATION:

; APPLICANT: HUEBNER, Robert C.

; APPLICANT: NORMAN, Jon A.

; APPLICANT: LIANG, Xiaowu

; APPLICANT: CARNER, Kristin R.

; APPLICANT: BARBOUR, Alan G.

; APPLICANT: LUKE, Catherine J.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford, P.C.

; STREET: 530 Fifth Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/663,998

; FILING DATE: 06-JUN-1996

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Kowalski, Thomas J.

; REGISTRATION NUMBER: 32,147

```

: REFERENCE/DOCKET NUMBER: 454312-2440
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-840-3333
:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5682 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
: US-08-663-998-4

```

Query Match 56.4%; Score 18.6; DB 2; Length 5682;  
Best Local Similarity 72.7%; Pred. No. 22;  
Matches 24; Conservative 0; Mismatches 9; Indels 0

**Qy**      1    tcgggatccttttatggttgattacactgactgc    33  
         || | | | | | | | | | | | | | | | |  
**Dd**     2454   TCTGGATCCTTAAAGGTTTTTTTGACTTTCTGC    2422

```

RESULT 15
US-07-854-596B-50
; Sequence 50, Application US/07854596B
; Patent No. 5434073
; GENERAL INFORMATION:
; APPLICANT: Dawson, Keith M
; APPLICANT: Hunter, Michael G
; APPLICANT: Czaplowski, Lloyd G
; TITLE OF INVENTION: Proteins and nucleic acids
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. John J. McDonnell
; STREET: Ten South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/854,596B
; FILING DATE: 03-JUN-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 92,337
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..41
; OTHER INFORMATION: /note= "oligonucleotide BB25
US-07-854-596B-50

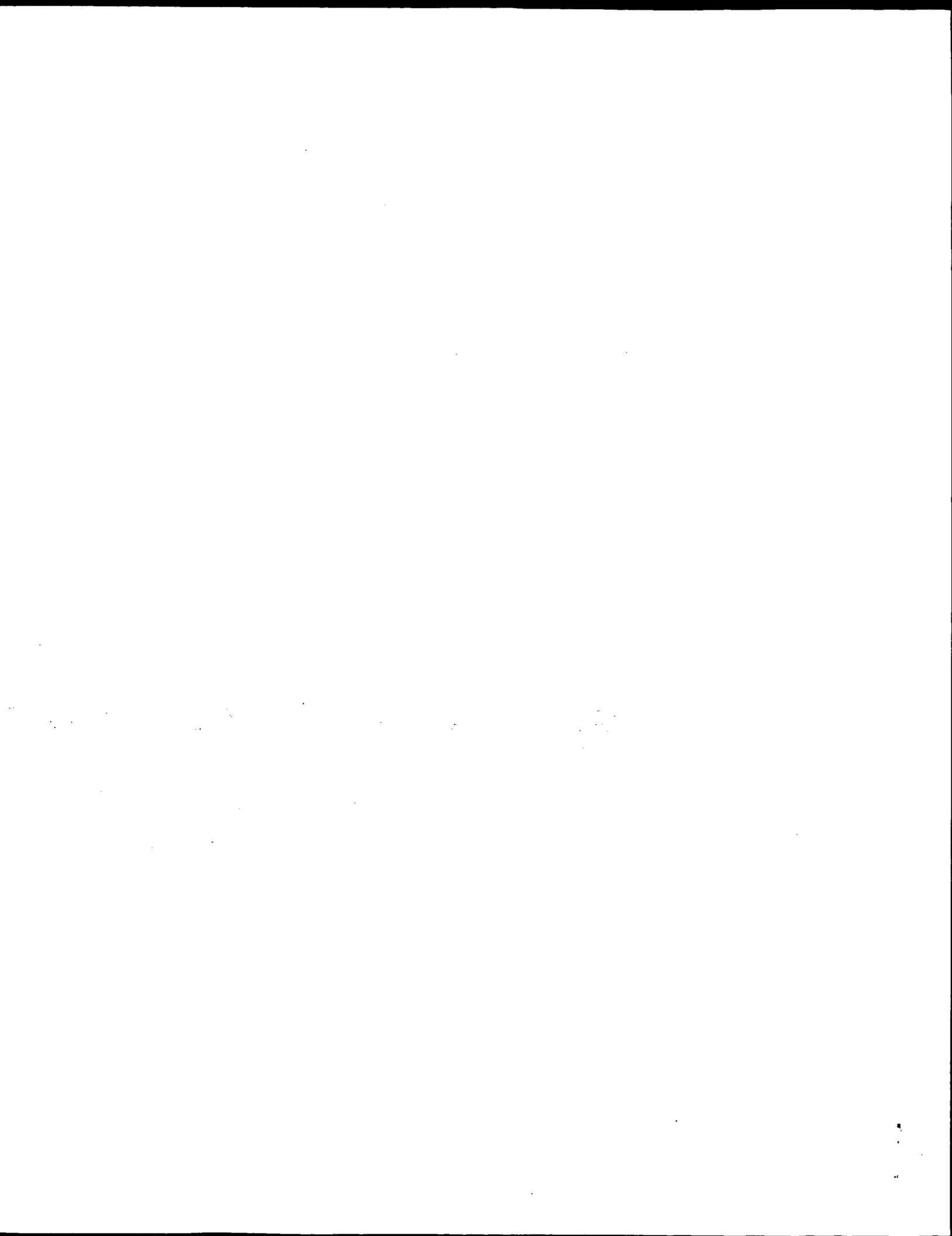
```

Query Match 55.8%; Score 18.4; DB 1; Length 41;  
Best Local Similarity 95.0%; Pred. NO. 9.8;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 13 atggttggttacactgactg 32  
|||||  
Db 13 ATGGTTGTTACACCGACTG 32

Search completed: July 15, 2002, 22:31:51  
Job time: 18477 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:49:01 ; Search time 6165.88 Seconds  
(without alignments)  
72.236 Million cell updates/sec

Title: US-10-053-641-5

Perfect score: 33

Sequence: 1 tcgggaccccttatggtgtttacactgactgc 33

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estcov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	23.2	70.3	954	12	CNS04PFY	AL301255 Tetraodon
C 2	21.4	64.8	976	12	CNS002KR	AL097797 Drosophila
C 3	21	63.6	286	10	BF329572	BF329572 CMO-BN027
C 4	20.8	63.0	638	10	BJ069739	BJ069739 BJO69739
C 5	20.8	63.0	673	10	BJ086834	BJ086834 BJO86834
C 6	20.6	62.4	1090	12	CNS05M72	AL343703 Tetraodon
C 7	20.4	61.8	611	12	AQ271191	AQ271191 nbxb0024K
C 8	20.4	61.8	644	10	BM113581	BM113581 EST561117
C 9	20.4	61.8	654	10	BE472919	BE472919 EST417772
C 10	20.4	61.8	668	12	AG139789	AG139789 Pan trogl
C 11	20.4	61.8	702	10	BI436222	BI436222 EST538983
C 12	20.4	61.8	726	10	BI176510	BI176510 EST517455
C 13	20.4	61.8	739	12	AQ688874	AQ688874 nbxb0078B
C 14	20.4	61.8	859	10	BF030472	BF030472 601558578
C 15	20.4	61.8	1116	12	CNS06SZ4	AL413894 T3 end of
C 16	20.2	61.2	267	9	AW193036	AW193036 xl69d03.x
C 17	20.2	61.2	268	10	BE302044	BE302044 bb80d03.x

C 18	20.2	61.2	314	9	AI377445	AI377445 tc23f06.x
C 19	20.2	61.2	328	9	AI632897	AI632897 CX35B02.x
C 20	20.2	61.2	367	10	N62121	N62121 Yz61g08.s1
C 21	20.2	61.2	381	10	N49233	N49233 Yy83b09.s1
C 22	20.2	61.2	456	10	M89275	M89275 CEL20D3 Chr
C 23	20.2	61.2	503	10	BE459104	BE459104 EST414396
C 24	20.2	61.2	510	9	AW930328	AW930328 EST340881
C 25	20.2	61.2	511	12	AZ148946	AZ148946 SP_0003.A
C 26	20.2	61.2	544	10	BG132676	BG132676 EST465568
C 27	20.2	61.2	565	10	BF098546	BF098546 EST429067
C 28	20.2	61.2	586	12	AZ176049	AZ176049 SP_0140.B
C 29	20.2	61.2	600	9	AI780646	AI780646 EST261621
C 30	20.2	61.2	639	10	BI925006	BI925006 EST544895
C 31	20.2	61.2	672	9	BB326689	BB326689 BB326689
C 32	20.2	61.2	912	12	AZ202576	AZ202576 SP_0056.B
C 33	20	60.6	475	12	AQ270392	AQ270392 HS_2052.B
C 34	20	60.6	707	12	BH016617	BH016617 TDGCO01TH
C 35	20	60.6	725	12	BH016891	BH016891 TDGCO096TH
C 36	19.8	60.0	126	9	AW993381	AW993381 RC2-BN003
C 37	19.8	60.0	174	12	AZ111323	AZ111323 RBC1-23-4
C 38	19.8	60.0	342	9	BB232269	BB232269 BB232269
C 39	19.8	60.0	404	9	AI342666	AI342666 qo31b09.x
C 40	19.8	60.0	476	10	N63009	N63009 Yy99b11.s1
C 41	19.8	60.0	495	10	BG149467	BG149467 nad29b06.
C 42	19.8	60.0	571	9	AI700684	AI700684 we39c12.x
C 43	19.8	60.0	608	12	AQ855887	AQ855887 CP61626B
C 44	19.8	60.0	873	10	BG678841	BG678841 602634737
C 45	19.8	60.0	953	12	CNS01VMO	AL169305 Tetraodon

#### ALIGNMENTS

#### RESULT 1

#### CNS04PFY/C

#### LOCUS

#### DEFINITION

#### ACCSSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### COMMENT

#### FEATURES

#### source

CNS04PFY 954 bp DNA linear GSS 24-MAY-2000  
Tetraodon nigroviridis genome survey sequence PUC-ori end of clone  
127M24 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

AL301255 GI:8178295  
AL301255.1 G1:8178295  
GSS: genome survey sequence.  
Tetraodon nigroviridis.  
Tetraodon nigroviridis.

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.

1 (bases 1 to 954)  
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,  
Bounneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and  
Weissenbach, J.

Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished

2 (bases 1 to 954)  
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bounneau, L., Fisher, C.,  
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,  
Saurin, W. and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished

3 (bases 1 to 954)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.

Location/Qualifiers  
1. .954  
/organism="Tetraodon nigroviridis"

KEYWORDS  
SOURCE  
ORGANISM

EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&st2=CM0-BN0270-  
240800-396-all&t3=2000-08-24&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 26  
High quality sequence stop: 286.

FEATURES  
source

1..286  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0270"  
/dev\_stage="Adult"  
/note="Organ: breast normal; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
80 a 58 c 73 g 75 t

BASE COUNT  
ORIGIN

Query Match 63.6%; Score 21; DB 10; Length 286;  
Best Local Similarity 82.8%; Pred. No. 4.6e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 gggatcccttattggtgtttacactgact 31  
||| ||||| ||||| ||||| |||||  
Db 143 GGAATTTTATGGCTTTTACACTGACT 171

RESULT 4  
BJ069739/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BJ069739 638 bp mRNA linear EST 11-DEC-2001  
BJ069739 NIBB Mochii normalized Xenopus tailbud library Xenopus  
laevis cDNA clone XL054920 5', mRNA sequence.  
BJ069739  
BJ069739.1 GI:17498099  
EST.  
African clawed frog.  
Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
Xenopodinae; Xenopus.  
1 (bases 1 to 638)  
Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara  
Y.

TITLE  
JOURNAL  
COMMENT  
Expressed genes in *X. laevis* embryo  
Unpublished (2001)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

## FEATURES

source  
1..638  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="XL054920"  
/clone\_lib="NIBB Mochii normalized Xenopus tailbud library"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 25"

BASE COUNT 173 a 153 c 157 g 155 t  
ORIGIN

Query Match 63.0%; Score 20.8; DB 10; Length 638;  
Best Local Similarity 78.1%; Pred. NO. 5.8e+02;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 tcggatcctttatggtgtttacactgactg 32  
||||| ||| ||| ||| ||| ||| ||| |||

Db 628 TTGGGATCTTTCTGGATATTTTCACCTGCCTG 597

## RESULT 5

LOCUS BJ086834 673 bp mRNA linear EST 12-DEC-2001  
DEFINITION BJ086834 NIBB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone XL054920 3', mRNA sequence.  
ACCESSION BJ086834  
VERSION BJ086834.1 GI:17583393  
KEYWORDS EST.  
SOURCE African clawed frog.  
ORGANISM *Xenopus laevis*

REFERENCE 1 (bases 1 to 673)  
AUTHORS Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.  
COMMENT  
Expressed genes in *X. laevis* embryo  
Unpublished (2001)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

## FEATURES

source  
1..673  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="XL054920"  
/clone\_lib="NIBB Mochii normalized Xenopus tailbud library"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 25"

BASE COUNT 194 a 151 c 154 g 172 t  
ORIGIN

Query Match 63.0%; Score 20.8; DB 10; Length 673;  
Best Local Similarity 78.1%; Pred. NO. 5.8e+02;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 tcggatcctttatggtgtttacactgactg 32  
||||| ||| ||| ||| ||| ||| ||| |||

Db 419 TTGGGATCTTTCTGGATATTTTCACCTGCCTG 450

## RESULT 6

LOCUS CNS05W72 1090 bp DNA linear GSS 26-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone 039D07 of library A from Tetraodon nigroviridis, genomic survey sequence.  
ACCESSION AL343703  
VERSION AL343703.1 GI:8237473  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
1 (bases 1 to 1090)  
Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bounau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.  
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 1090)  
Roest-Crolius, H., Jaillon, O., Dasilva, C., Bounau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.  
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 1090)  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
Source  
1..1090  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="039D07"  
/clone\_lib="A"  
/notes="Genoscope sequence ID : COAA039CB04A1-end : T3"  
BASE COUNT 367 a 192 c 230 g 288 t  
ORIGIN

Query Match 62.4%; Score 20.6; DB 12; Length 1090;  
Best Local Similarity 79.3%; Pred. No. 7.2e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 4 ggatcctttatggtgtttacactgactg 32  
||||| ||| ||| ||| ||| ||| ||| |||

Db 868 GGATCAATATGTTTGTATACACAG 840

## RESULT 7

LOCUS AQ271191 611 bp DNA linear GSS 03-NOV-1998  
DEFINITION nbxb0024K04r CUGI Rice BAC Library Oryza sativa genomic clone nbxb0024K04r, DNA sequence.  
ACCESSION AQ271191  
VERSION AQ271191.1 GI:3824506  
KEYWORDS GSS.  
SOURCE Oryza sativa  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoidae; Orzyeae; Oryza.

1 (bases 1 to 611)  
Wing,R.A. and Dean,R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu

Seq primer: GGAACAGCTATGACCATG  
Class: BAC ends  
High quality sequence start: 9  
High quality sequence stop: 345.

Location/Qualifiers  
1. .611

/organism="Oryza sativa"  
/strain="Japonica"

/cultivar="Nipponbare"  
/db\_xref="taxon:4530"

/clone="nbxb0024K04r"  
/clone\_lib="CUGI Rice BAC Library"

/tissue\_type="Leaf"  
/lab\_host="E. coli DH10B"

/note="Vector: pBluescript SK(-); Site\_1: HindIII; Site\_2:  
HindIII; Rice is one of two most popular grains in the  
world. Half of the world population especially those  
inhabiting highly populated areas of the humid tropics  
and subtropics, rely on rice as their primary source of  
carbohydrate. Monocotyledonous rice is a diploid plant  
(2n=24) with a haploid genome equivalent of 431 Mbp  
(Arumuganathan and Earle, 1991). The relatively small  
genome of rice, three times larger than that of  
Arabidopsis, makes it suitable for genomic studies. In  
order to facilitate positional cloning, physical mapping  
and genome sequencing of rice, we have constructed a BAC  
library from Oryza sativa, Nipponbare variety. The  
library contains 36,864 clones with an average insert size  
of 128.5 Kb providing 10.9 haploid genome equivalents. The  
deep coverage allows the isolation a particular sequence  
with a probability of 99.9 %. Two high density filters,  
each containing 18,432 clones (doubly spotted), represent  
the whole library for colony screening."

174 a 132 c 120 g 184 t 1 others

BASE COUNT 174 a 132 c 120 g 184 t 1 others

ORIGIN

Query Match 61.8%; Score 20.4; DB 12; Length 611;  
Best Local Similarity 80.0%; Pred. No. 8.2e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 tcgggatcccttattggtgtttacactgac 30  
||||||| ||| ||||| |||

Db 97 TTGGATCCCTTGGGCTGTTTACATGGC 68  
||||||| ||| ||||| |||

RESULT 8

BM113581

LOCUS

DEFINITION

EST561117 potato roots Solanum tuberosum cDNA clone CPRO18D22 5'

end, mRNA sequence.

ACCESSION

BM113581

VERSION

BM113581.1 GI:17076629

KEYWORDS

EST.

SOURCE

Solanum tuberosum

potato.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 644)

REFERENCE

1 (bases 1 to 611)

Wing,R.A. and Dean,R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

# AUTHORS

van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S.,  
Utterback,T., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C.,  
Tanksley,S. and Baker,B.

Generation of ESTs from potato roots

Unpublished (2001)

Contact: Research Genetics, Libraries Division

Tel: 1-800-711-6195

Email: cdna@resgen.com

For clone info: please contact Research Genetics, Libraries

Division tel 1-800-711-6195, email cdna@resgen.com

Seq primer: T3.

Location/Qualifiers

1. .644

/organism="Solanum tuberosum"

/cultivar="Kennebec"

/db\_xref="taxon:4113"

/clone="cPRO18D22"

/clone\_lib="potato roots"

/tissue\_type="roots"

/dev\_stage="in vitro grown stem cuttings"

/lab\_host="SOLR"

/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Cornell University, Tanksley lab;  
sequencing: The Institute for Genomic Research. Roots were  
isolated from in vitro grown stem cuttings on CM medium.  
Roots were isolated two weeks after placing the stem  
cuttings from in vitro grown plants on medium."

153 a 134 c 147 g 210 t

BASE COUNT 153 a 134 c 147 g 210 t

ORIGIN

Query Match 61.8%; Score 20.4; DB 10; Length 644;  
Best Local Similarity 80.0%; Pred. No. 8.3e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 4 ggatcttattggtgtttacactgactgc 33  
||||| ||| ||| ||| ||| ||| ||| |||

Db 557 GGAAGTTTGCTGCTGTTCACACTGCTC 586  
||||| ||| ||| ||| ||| ||| ||| |||

RESULT 9

BE472919

LOCUS

DEFINITION

EST417772 potato stolon, Cornell University Solanum tuberosum cDNA

clone cSTA3506, mRNA sequence.

ACCESSION

BE472919

VERSION

BE472919.1 GI:9563410

KEYWORDS

EST.

SOURCE

potato.

ORGANISM

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 654)

REFERENCE

AUTHORS

van der Hoeven,R., Beizerides,J., Bachem,C., Horvath,B., Visser,R.,  
Holt,I.E., Liang,F., Hansen,T.S., Utterback,T., Bowman,C.L., Doan  
B., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker  
B.

Generation of ESTs from potato swelling stolons

Unpublished (1999)

Contact: Research Genetics, Libraries Division

Tel: 1-800-711-6195

Email: cdna@resgen.com

5 prime sequence.

Location/Qualifiers

1. .654

/organism="Solanum tuberosum"

/cultivar="Bintje"

/db\_xref="taxon:4113"

/clone="cSTA3506"

/clone\_lib="potato stolon, Cornell University"

/tissue\_type="axillary buds of stem explants, swelling  
stolons"

source



Best Local Similarity 80.0%; Pred. No. 8.3e+02;									
Matches 24;		Conservative 0;		Mismatches 6;		Indels 0;		Gaps 0;	
Qy	4	ggatcccttattggtgtttacactgactgc	33						
Db	282	GGATCCCTTCTCTGTGTTCACACAGCCG	311						
RESULT 11									
BI436222									
LOCUS		BI436222		702 bp		mRNA		linear	
DEFINITION		EST538983 cSTE Solanum tuberosum cDNA clone cSTE22C11 5' sequence, mRNA sequence.							

KEYWORDS EST. SOURCE ORGANISM

potato.  
Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
(bases 1 to 702)

REFERENCE

van der Hoeven, R., Bezzarides, J., Bachem, C., Visser, R., Cho, J.,

**TITLE** Generation of ESTs from in vitro grown microtubers  
**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Cathy Ronning  
 The Institute for Genomic Research  
 For clone info: please contact Research Genetics, Libraries  
 Division tel 1-800-711-6195, email cdna@resgen.com  
 Seq primer: M13F-R.  
**FEATURES** Location/Qualifiers  
 source 1..702

```

1. : 702
/organism="Solanum tuberosum"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="cSTe22C11"
/clone_lib="cSTe"
/tissue_type="axillary buds of stem explants; growing
sink-tubers"
/dev_stage="7, 8 and 10 days"
/lab_host="SOLR"
/note="Vector: Bluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Tissue supplied by Christian Bachem and Richard
Vlaiser (Department of Plant Breeding, Wageningen
University, The Netherlands). The cSTa libraries will
attempt to capture the induction and initiation/initial
growth of the tuber in an in vitro system as described in
Bachem et al. (Plant Journal, 1996). Small microtubers
develop from axillary buds attached to stem explants when
placed on a high sucrose medium (10%). Visible
morphological changes occur synchronously at day five in
the axillary buds. The first library, cSTa (1-20) consists

```

[illegible]

RESULT 12



# TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LICM498 row: 1 column: 19  
High quality sequence stop: 111.

## FEATURES source

Location/Qualifiers  
1. 859  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3828258"  
/clone\_lib="NIH\_MGC\_58"  
/tissue\_type="hypertrophoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-ATTCTAGAGCGGCGCGGACATG-dT(30)BN-3',  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.35 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."  
BASE COUNT 270 a 315 c 177 g 97 t  
ORIGIN

Query Match 61.8%; Score 20.4; DB 10; Length 859;  
Best Local Similarity 80.0%; Pred. No. 8.4e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 gggatccttattggtgtttacactgactg 32  
||| ||||| ||| ||| ||||| ||  
DB 78 GGGCTCTTATTGTTTATAAAGTGAATG 107

## RESULT 15 CNS06S24/c

LOCUS  
DEFINITION T3 end of clone AW0AA027F06 of library AW0AA from strain CLIB 89 of Yarrowia lipolytica, genomic survey sequence.  
ACCESSION AL413894  
VERSION AL413894.1 GI:12186382  
KEYWORDS GSS.

## SOURCE ORGANISM

Yarrowia lipolytica.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Dipodasaceae; Yarrowia.  
1 (bases 1 to 1116)

## REFERENCE AUTHORS

Casaregola, S., Neuveglise, C., Lepingle, A., Bon, E., Feynerol, C.,  
Artiguenave, F., Wincker, P. and Gaillardin, C.  
Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia  
lipolytica  
FEBS Lett. 487 (1), 95-100 (2000)  
20584727

## REFERENCE AUTHORS

2 (bases 1 to 1116)  
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,  
Bollotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,  
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Liorente, B.,  
Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,  
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
Wincker, P. and Weissenbach, J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies

# JOURNAL MEDLINE REFERENCE AUTHORS TITLE JOURNAL

FEBS Lett. 487 (1), 3-12 (2000)  
20584711  
3 (bases 1 to 1116)  
Genoscope.  
Direct Submission

## COMMENT

Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
seqref@genoscope.cns.fr - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

## FEATURES source

Location/Qualifiers  
1. 1116  
/organism="Yarrowia lipolytica"  
/strain="CLIB 89"  
/db\_xref="taxon:4952"  
/clone="AW0AA027F06"  
/clone\_lib="AW0AA"  
/note="end : T3"  
BASE COUNT 325 a 224 c 272 g 292 t 3 others  
ORIGIN

Query Match 61.8%; Score 20.4; DB 12; Length 1116;  
Best Local Similarity 80.0%; Pred. No. 8.6e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 gggatccttattggtgtttacactgactg 32  
||| ||||| ||||| ||||| |||  
DB 870 GGTCTCTAGATGGTGTTCACACTCATG 841

Search completed: July 15, 2002, 21:49:08  
Job time: 18899 sec







```

ORGANISM      Unknown.
Unclassified.
REFERENCE      1 (bases 1 to 59)
AUTHORS        Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
                Hoeffken,H.Wolfgang.
TITLE          Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL        Patent: US 5663141-A 9 02-SEP-1997;
FEATURES       Location/Qualifiers
source         1..59
BASE COUNT     20 a 14 c 15 g 10 t
ORIGIN

Query Match      81.9%; Score 25.4; DB 6; Length 59;
Best Local Similarity 96.3%; Pred. No. 0.14;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gctctagagccaggatgcatgtt 27
    ||| ||||| ||||| ||||| ||||| |||||
Db 27 GCCTTTAGAGCCAGGATGCAATTGTT 1

RESULT 7
I63548/c
LOCUS          I63548
DEFINITION     Sequence 11 from patent US 5663141.
ACCESSION      I63548
VERSION        I63548.1 GI:2481121
KEYWORDS       Unknown.
SOURCE         Unknown.
ORGANISM       Unclassified.
1 (bases 1 to 59)
AUTHORS        Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
                Hoeffken,H.Wolfgang.
TITLE          Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL        Patent: US 5663141-A 11 02-SEP-1997;
FEATURES       Location/Qualifiers
source         1..59
BASE COUNT     18 a 14 c 16 g 11 t
ORIGIN

Query Match      81.9%; Score 25.4; DB 6; Length 59;
Best Local Similarity 96.3%; Pred. No. 0.14;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gctctagagccaggatgcatgtt 27
    ||| ||||| ||||| ||||| ||||| |||||
Db 27 GCCTTTAGAGCCAGGATGCAATTGTT 1

RESULT 8
I41183/c
LOCUS          I41183
DEFINITION     Sequence 3 from patent US 5624822.
ACCESSION      I41183
VERSION        I41183.1 GI:2081773
KEYWORDS       Unknown.
SOURCE         Unknown.
ORGANISM       Unclassified.
1 (bases 1 to 129)
AUTHORS        Koerwer,W.
TITLE          Hirudin fusion proteins and preparation of hirudin
JOURNAL        Patent: US 5624822-A 3 29-APR-1997;
FEATURES       Location/Qualifiers
source         1..129
BASE COUNT     36 a 35 c 29 g 29 t
ORIGIN

```

```

Query Match      80.6%; Score 25; DB 6; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 agagccagagatgcatgttggcc 31
    ||| ||||| ||||| ||||| ||||| |||||
Db 129 AGAGCCAGGATGCAATTGTGCCC 105

RESULT 9
MMHC425018
LOCUS          MMHC425018
DEFINITION     Mus musculus major histocompatibility complex region NG27, NG28,
                RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein,
                BING1, Tapesin, RalGDS-like, KE2, BING4, beta 1.3-galactosyl
                transferase, and RPS18 genes, complete cds; Sacm21 gene, partial
                cds; and unknown gene.
ACCESSION      AF110520
VERSION        AF110520.1 GI:4050090
KEYWORDS       house mouse.
SOURCE         Mus musculus.
ORGANISM       Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 213245)
AUTHORS        Rowen,L., Qin,S., Madan,A., Loretz,C., Hall,J., James,R., Dors,M.,
                Shaffer,T., Abbasi,N., Ratcliffe,A., Dickhoff,R., Lasky,S. and
                Hood,L.
TITLE          Sequence of the mouse major histocompatibility complex class II
                region
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 213245)
AUTHORS        Rowen,L.
TITLE          Direct Submission
JOURNAL        Submitted (03-DEC-1998) Department of Molecular Biotechnology, Box
                357730 University of Washington, Seattle, Washington 98195, USA
COMMENT        Sequencing methodology: high redundancy shotgun. Interspersed
                Repeats were identified with RepeatMasker (available from
                http://ftp.genome.washington.edu/RM/RepeatMasker.html). This clone
                overlaps clone 445d10, found in GenBankAccession Number AF100956.
                It is possible that this clone is aberrant, since it does not
                contain the SYNGAP Atpase found 5' of HSET in the human counterpart
                region (kinesin, in this sequence). It is not known yet whether
                this clone has deleted the SYNGAP gene or whether the 129 strain
                has deleted it.
FEATURES       Location/Qualifiers
source         1..213245
                /organism="Mus musculus"
                /strain="129"
                /db_xref="taxon:10090"
                /chromosome="17"
                /clone="BAC425018"
                /clone_lib="Genome Systems"
                /complement(422..504)
                /rpt_family="PB1D10"
                /complement(1421..1479)
                /rpt_family="PB1D10"
                /complement(1421..1553)
                /rpt_family="MMB1F"
                /complement(1514..1642)
                /rpt_family="B4A"
                /complement(1568..1642)
                /rpt_family="MUSID5"
                /complement(1895..2375)
                /rpt_family="LINE2"
                /complement(2573..2713)
                /rpt_family="MMB1F"
                /complement(2719..2793)
                /rpt_family="MUSID4"
                /complement(2951..3048)
                /rpt_family="PB1D10"
                /complement(3048..3129)
                /rpt_family="PB1D10"
                /complement(3129..3210)
                /rpt_family="PB1D10"
                /complement(3210..3291)
                /rpt_family="PB1D10"
                /complement(3291..3372)
                /rpt_family="PB1D10"
                /complement(3372..3453)
                /rpt_family="PB1D10"
                /complement(3453..3534)
                /rpt_family="PB1D10"
                /complement(3534..3615)
                /rpt_family="PB1D10"
                /complement(3615..3696)
                /rpt_family="PB1D10"
                /complement(3696..3777)
                /rpt_family="PB1D10"
                /complement(3777..3858)
                /rpt_family="PB1D10"
                /complement(3858..3939)
                /rpt_family="PB1D10"
                /complement(3939..4020)
                /rpt_family="PB1D10"
                /complement(4020..4101)
                /rpt_family="PB1D10"
                /complement(4101..4182)
                /rpt_family="PB1D10"
                /complement(4182..4263)
                /rpt_family="PB1D10"
                /complement(4263..4344)
                /rpt_family="PB1D10"
                /complement(4344..4425)
                /rpt_family="PB1D10"
                /complement(4425..4506)
                /rpt_family="PB1D10"
                /complement(4506..4587)
                /rpt_family="PB1D10"
                /complement(4587..4668)
                /rpt_family="PB1D10"
                /complement(4668..4749)
                /rpt_family="PB1D10"
                /complement(4749..4830)
                /rpt_family="PB1D10"
                /complement(4830..4911)
                /rpt_family="PB1D10"
                /complement(4911..4992)
                /rpt_family="PB1D10"
                /complement(4992..5073)
                /rpt_family="PB1D10"
                /complement(5073..5154)
                /rpt_family="PB1D10"
                /complement(5154..5235)
                /rpt_family="PB1D10"
                /complement(5235..5316)
                /rpt_family="PB1D10"
                /complement(5316..5397)
                /rpt_family="PB1D10"
                /complement(5397..5478)
                /rpt_family="PB1D10"
                /complement(5478..5559)
                /rpt_family="PB1D10"
                /complement(5559..5640)
                /rpt_family="PB1D10"
                /complement(5640..5721)
                /rpt_family="PB1D10"
                /complement(5721..5802)
                /rpt_family="PB1D10"
                /complement(5802..5883)
                /rpt_family="PB1D10"
                /complement(5883..5964)
                /rpt_family="PB1D10"
                /complement(5964..6045)
                /rpt_family="PB1D10"
                /complement(6045..6126)
                /rpt_family="PB1D10"
                /complement(6126..6207)
                /rpt_family="PB1D10"
                /complement(6207..6288)
                /rpt_family="PB1D10"
                /complement(6288..6369)
                /rpt_family="PB1D10"
                /complement(6369..6450)
                /rpt_family="PB1D10"
                /complement(6450..6531)
                /rpt_family="PB1D10"
                /complement(6531..6612)
                /rpt_family="PB1D10"
                /complement(6612..6693)
                /rpt_family="PB1D10"
                /complement(6693..6774)
                /rpt_family="PB1D10"
                /complement(6774..6855)
                /rpt_family="PB1D10"
                /complement(6855..6936)
                /rpt_family="PB1D10"
                /complement(6936..7017)
                /rpt_family="PB1D10"
                /complement(7017..7098)
                /rpt_family="PB1D10"
                /complement(7098..7179)
                /rpt_family="PB1D10"
                /complement(7179..7260)
                /rpt_family="PB1D10"
                /complement(7260..7341)
                /rpt_family="PB1D10"
                /complement(7341..7422)
                /rpt_family="PB1D10"
                /complement(7422..7503)
                /rpt_family="PB1D10"
                /complement(7503..7584)
                /rpt_family="PB1D10"
                /complement(7584..7665)
                /rpt_family="PB1D10"
                /complement(7665..7746)
                /rpt_family="PB1D10"
                /complement(7746..7827)
                /rpt_family="PB1D10"
                /complement(7827..7908)
                /rpt_family="PB1D10"
                /complement(7908..7989)
                /rpt_family="PB1D10"
                /complement(7989..8070)
                /rpt_family="PB1D10"
                /complement(8070..8151)
                /rpt_family="PB1D10"
                /complement(8151..8232)
                /rpt_family="PB1D10"
                /complement(8232..8313)
                /rpt_family="PB1D10"
                /complement(8313..8394)
                /rpt_family="PB1D10"
                /complement(8394..8475)
                /rpt_family="PB1D10"
                /complement(8475..8556)
                /rpt_family="PB1D10"
                /complement(8556..8637)
                /rpt_family="PB1D10"
                /complement(8637..8718)
                /rpt_family="PB1D10"
                /complement(8718..8799)
                /rpt_family="PB1D10"
                /complement(8799..8880)
                /rpt_family="PB1D10"
                /complement(8880..8961)
                /rpt_family="PB1D10"
                /complement(8961..9042)
                /rpt_family="PB1D10"
                /complement(9042..9123)
                /rpt_family="PB1D10"
                /complement(9123..9204)
                /rpt_family="PB1D10"
                /complement(9204..9285)
                /rpt_family="PB1D10"
                /complement(9285..9366)
                /rpt_family="PB1D10"
                /complement(9366..9447)
                /rpt_family="PB1D10"
                /complement(9447..9528)
                /rpt_family="PB1D10"
                /complement(9528..9609)
                /rpt_family="PB1D10"
                /complement(9609..9690)
                /rpt_family="PB1D10"
                /complement(9690..9771)
                /rpt_family="PB1D10"
                /complement(9771..9852)
                /rpt_family="PB1D10"
                /complement(9852..9933)
                /rpt_family="PB1D10"
                /complement(9933..10014)
                /rpt_family="PB1D10"
                /complement(10014..10095)
                /rpt_family="PB1D10"
                /complement(10095..10176)
                /rpt_family="PB1D10"
                /complement(10176..10257)
                /rpt_family="PB1D10"
                /complement(10257..10338)
                /rpt_family="PB1D10"
                /complement(10338..10419)
                /rpt_family="PB1D10"
                /complement(10419..10500)
                /rpt_family="PB1D10"
                /complement(10500..10581)
                /rpt_family="PB1D10"
                /complement(10581..10662)
                /rpt_family="PB1D10"
                /complement(10662..10743)
                /rpt_family="PB1D10"
                /complement(10743..10824)
                /rpt_family="PB1D10"
                /complement(10824..10905)
                /rpt_family="PB1D10"
                /complement(10905..10986)
                /rpt_family="PB1D10"
                /complement(10986..11067)
                /rpt_family="PB1D10"
                /complement(11067..11148)
                /rpt_family="PB1D10"
                /complement(11148..11229)
                /rpt_family="PB1D10"
                /complement(11229..11310)
                /rpt_family="PB1D10"
                /complement(11310..11391)
                /rpt_family="PB1D10"
                /complement(11391..11472)
                /rpt_family="PB1D10"
                /complement(11472..11553)
                /rpt_family="PB1D10"
                /complement(11553..11634)
                /rpt_family="PB1D10"
                /complement(11634..11715)
                /rpt_family="PB1D10"
                /complement(11715..11796)
                /rpt_family="PB1D10"
                /complement(11796..11877)
                /rpt_family="PB1D10"
                /complement(11877..11958)
                /rpt_family="PB1D10"
                /complement(11958..12039)
                /rpt_family="PB1D10"
                /complement(12039..12120)
                /rpt_family="PB1D10"
                /complement(12120..12201)
                /rpt_family="PB1D10"
                /complement(12201..12282)
                /rpt_family="PB1D10"
                /complement(12282..12363)
                /rpt_family="PB1D10"
                /complement(12363..12444)
                /rpt_family="PB1D10"
                /complement(12444..12525)
                /rpt_family="PB1D10"
                /complement(12525..12606)
                /rpt_family="PB1D10"
                /complement(12606..12687)
                /rpt_family="PB1D10"
                /complement(12687..12768)
                /rpt_family="PB1D10"
                /complement(12768..12849)
                /rpt_family="PB1D10"
                /complement(12849..12930)
                /rpt_family="PB1D10"
                /complement(12930..13011)
                /rpt_family="PB1D10"
                /complement(13011..13092)
                /rpt_family="PB1D10"
                /complement(13092..13173)
                /rpt_family="PB1D10"
                /complement(13173..13254)
                /rpt_family="PB1D10"
                /complement(13254..13335)
                /rpt_family="PB1D10"
                /complement(13335..13416)
                /rpt_family="PB1D10"
                /complement(13416..13497)
                /rpt_family="PB1D10"
                /complement(13497..13578)
                /rpt_family="PB1D10"
                /complement(13578..13659)
                /rpt_family="PB1D10"
                /complement(13659..13740)
                /rpt_family="PB1D10"
                /complement(13740..13821)
                /rpt_family="PB1D10"
                /complement(13821..13902)
                /rpt_family="PB1D10"
                /complement(13902..13983)
                /rpt_family="PB1D10"
                /complement(13983..14064)
                /rpt_family="PB1D10"
                /complement(14064..14145)
                /rpt_family="PB1D10"
                /complement(14145..14226)
                /rpt_family="PB1D10"
                /complement(14226..14307)
                /rpt_family="PB1D10"
                /complement(14307..14388)
                /rpt_family="PB1D10"
                /complement(14388..14469)
                /rpt_family="PB1D10"
                /complement(14469..14550)
                /rpt_family="PB1D10"
                /complement(14550..14631)
                /rpt_family="PB1D10"
                /complement(14631..14712)
                /rpt_family="PB1D10"
                /complement(14712..14793)
                /rpt_family="PB1D10"
                /complement(14793..14874)
                /rpt_family="PB1D10"
                /complement(14874..14955)
                /rpt_family="PB1D10"
                /complement(14955..15036)
                /rpt_family="PB1D10"
                /complement(15036..15117)
                /rpt_family="PB1D10"
                /complement(15117..15198)
                /rpt_family="PB1D10"
                /complement(15198..15279)
                /rpt_family="PB1D10"
                /complement(15279..15360)
                /rpt_family="PB1D10"
                /complement(15360..15441)
                /rpt_family="PB1D10"
                /complement(15441..15522)
                /rpt_family="PB1D10"
                /complement(15522..15603)
                /rpt_family="PB1D10"
                /complement(15603..15684)
                /rpt_family="PB1D10"
                /complement(15684..15765)
                /rpt_family="PB1D10"
                /complement(15765..15846)
                /rpt_family="PB1D10"
                /complement(15846..15927)
                /rpt_family="PB1D10"
                /complement(15927..16008)
                /rpt_family="PB1D10"
                /complement(16008..16089)
                /rpt_family="PB1D10"
                /complement(16089..16170)
                /rpt_family="PB1D10"
                /complement(16170..16251)
                /rpt_family="PB1D10"
                /complement(16251..16332)
                /rpt_family="PB1D10"
                /complement(16332..16413)
                /rpt_family="PB1D10"
                /complement(16413..16494)
                /rpt_family="PB1D10"
                /complement(16494..16575)
                /rpt_family="PB1D10"
                /complement(16575..16656)
                /rpt_family="PB1D10"
                /complement(16656..16737)
                /rpt_family="PB1D10"
                /complement(16737..16818)
                /rpt_family="PB1D10"
                /complement(16818..16899)
                /rpt_family="PB1D10"
                /complement(16899..16980)
                /rpt_family="PB1D10"
                /complement(16980..17061)
                /rpt_family="PB1D10"
                /complement(17061..17142)
                /rpt_family="PB1D10"
                /complement(17142..17223)
                /rpt_family="PB1D10"
                /complement(17223..17304)
                /rpt_family="PB1D10"
                /complement(17304..17385)
                /rpt_family="PB1D10"
                /complement(17385..17466)
                /rpt_family="PB1D10"
                /complement(17466..17547)
                /rpt_family="PB1D10"
                /complement(17547..17628)
                /rpt_family="PB1D10"
                /complement(17628..17709)
                /rpt_family="PB1D10"
                /complement(17709..17790)
                /rpt_family="PB1D10"
                /complement(17790..17871)
                /rpt_family="PB1D10"
                /complement(17871..17952)
                /rpt_family="PB1D10"
                /complement(17952..18033)
                /rpt_family="PB1D10"
                /complement(18033..18114)
                /rpt_family="PB1D10"
                /complement(18114..18195)
                /rpt_family="PB1D10"
                /complement(18195..18276)
                /rpt_family="PB1D10"
                /complement(18276..18357)
                /rpt_family="PB1D10"
                /complement(18357..18438)
                /rpt_family="PB1D10"
                /complement(18438..18519)
                /rpt_family="PB1D10"
                /complement(18519..18600)
                /rpt_family="PB1D10"
                /complement(18600..18681)
                /rpt_family="PB1D10"
                /complement(18681..18762)
                /rpt_family="PB1D10"
                /complement(18762..18843)
                /rpt_family="PB1D10"
                /complement(18843..18924)
                /rpt_family="PB1D10"
                /complement(18924..19005)
                /rpt_family="PB1D10"
                /complement(19005..19086)
                /rpt_family="PB1D10"
                /complement(19086..19167)
                /rpt_family="PB1D10"
                /complement(19167..19248)
                /rpt_family="PB1D10"
                /complement(19248..19329)
                /rpt_family="PB1D10"
                /complement(19329..19410)
                /rpt_family="PB1D10"
                /complement(19410..19491)
                /rpt_family="PB1D10"
                /complement(19491..19572)
                /rpt_family="PB1D10"
                /complement(19572..19653)
                /rpt_family="PB1D10"
                /complement(19653..19734)
                /rpt_family="PB1D10"
                /complement(19734..19815)
                /rpt_family="PB1D10"
                /complement(19815..19896)
                /rpt_family="PB1D10"
                /complement(19896..19977)
                /rpt_family="PB1D10"
                /complement(19977..20058)
                /rpt_family="PB1D10"
                /complement(20058..20139)
                /rpt_family="PB1D10"
                /complement(20139..20220)
                /rpt_family="PB1D10"
                /complement(20220..20301)
                /rpt_family="PB1D10"
                /complement(20301..20382)
                /rpt_family="PB1D10"
                /complement(20382..20463)
                /rpt_family="PB1D10"
                /complement(20463..20544)
                /rpt_family="PB1D10"
                /complement(20544..20625)
                /rpt_family="PB1D10"
                /complement(20625..20706)
                /rpt_family="PB1D10"
                /complement(20706..20787)
                /rpt_family="PB1D10"
                /complement(20787..20868)
                /rpt_family="PB1D10"
                /complement(20868..20949)
                /rpt_family="PB1D10"
                /complement(20949..21030)
                /rpt_family="PB1D10"
                /complement(21030..21111)
                /rpt_family="PB1D10"
                /complement(21111..21192)
                /rpt_family="PB1D10"
                /complement(21192..21273)
                /rpt_family="PB1D10"
                /complement(21273..21354)
                /rpt_family="PB1D10"
                /complement(21354..21435)
                /rpt_family="PB1D10"
                /complement(21435..21516)
                /rpt_family="PB1D10"
                /complement(21516..21597)
                /rpt_family="PB1D10"
                /complement(21597..21678)
                /rpt_family="PB1D10"
                /complement(21678..21759)
                /rpt_family="PB1D10"
                /complement(21759..21840)
                /rpt_family="PB1D10"
                /complement(21840..21921)
                /rpt_family="PB1D10"
                /complement(21921..22002)
                /rpt_family="PB1D10"
                /complement(22002..22083)
                /rpt_family="PB1D10"
                /complement(22083..22164)
                /rpt_family="PB1D10"
                /complement(22164..22245)
                /rpt_family="PB1D10"
                /complement(22245..22326)
                /rpt_family="PB1D10"
                /complement(22326..22407)
                /rpt_family="PB1D10"
                /complement(22407..22488)
                /rpt_family="PB1D10"
                /complement(22488..22569)
                /rpt_family="PB1D10"
                /complement(22569..22650)
                /rpt_family="PB1D10"
                /complement(22650..22731)
                /rpt_family="PB1D10"
                /complement(22731..22812)
                /rpt_family="PB1D10"
                /complement(22812..22893)
                /rpt_family="PB1D10"
                /complement(22893..22974)
                /rpt_family="PB1D10"
                /complement(22974..23055)
                /rpt_family="PB1D10"
                /complement(23055..23136)
                /rpt_family="PB1D10"
                /complement(23136..23217)
                /rpt_family="PB1D10"
                /complement(23217..23298)
                /rpt_family="PB1D10"
                /complement(23298..23379)
                /rpt_family="PB1D10"
                /complement(23379..23460)
                /rpt_family="PB1D10"
                /complement(23460..23541)
                /rpt_family="PB1D10"
                /complement(23541..23622)
                /rpt_family="PB1D10"
                /complement(23622..23703)
                /rpt_family="PB1D10"
                /complement(23703..23784)
                /rpt_family="PB1D10"
                /complement(23784..23865)
                /rpt_family="PB1D10"
                /complement(23865..23946)
                /rpt_family="PB1D10"
                /complement(23946..24027)
                /rpt_family="PB1D10"
                /complement(24027..24108)
                /rpt_family="PB1D10"
                /complement(24108..24189)
                /rpt_family="PB1D10"
                /complement(24189..24270)
                /rpt_family="PB1D10"
                /complement(24270..24351)
                /rpt_family="PB1D10"
                /complement(24351..24432)
                /rpt_family="PB1D10"
                /complement(24432..24513)
                /rpt_family="PB1D10"
                /complement(24513..24594)
                /rpt_family="PB1D10"
                /complement(24594..24675)
                /rpt_family="PB1D10"
                /complement(24675..24756)
                /rpt_family="PB1D10"
                /complement(24756..24837)
                /rpt_family="PB1D10"
                /complement(24837..24918)
                /rpt_family="PB1D10"
                /complement(24918..25000)
                /rpt_family="PB1D10"
                /complement(25000..25081)
                /rpt_family="PB1D10"
                /complement(25081..25162)
                /rpt_family="PB1D10"
                /complement(25162..25243)
                /rpt_family="PB1D10"
                /complement(25243..25324)
                /rpt_family="PB1D10"
                /complement(25324..25405)
                /rpt_family="PB1D10"
                /complement(25405..25486)
                /rpt_family="PB1D10"
                /complement(25486..25567)
                /rpt_family="PB1D10"
                /complement(25567..25648)
                /rpt_family="PB1D10"
                /complement(25648..25729)
                /rpt_family="PB1D10"
                /complement(25729..25810)
                /rpt_family="PB1D10"
                /complement(25810..25891)
                /rpt_family="PB1D10"
                /complement(25891..25972)
                /rpt_family="PB1D10"
                /complement(25972..26053)
                /rpt_family="PB1D10"
                /complement(26053..26134)
                /rpt_family="PB1D10"
                /complement(26134..26215)
                /rpt_family="PB1D10"
                /complement(26215..26296)
                /rpt_family="PB1D10"
                /complement(26296..26377)
                /rpt_family="PB1D10"
                /complement(26377..26458)
                /rpt_family="PB1D10"
                /complement(26458..26539)
                /rpt_family="PB1D10"
                /complement(26539..26620)
                /rpt_family="PB1D10"
                /complement(26620..26701)
                /rpt_family="PB1D10"
                /complement(26701..26782)
                /rpt_family="PB1D10"
                /complement(26782..26863)
                /rpt_family="PB1D10"
                /complement(26863..26944)
                /rpt_family="PB1D10"
                /complement(26944..27025)
                /rpt_family="PB1D10"
                /complement(27025..27106)
                /rpt_family="PB1D10"
                /complement(27106..27187)
                /rpt_family="PB1D10"
                /complement(27187..27268)
                /rpt_family="PB1D10"
                /complement(27268..27349)
                /rpt_family="PB1D10"
                /complement(27349..27430)
                /rpt_family="PB1D10"
                /complement(27430..27511)
                /rpt_family="PB1D10"
                /complement(27511..27592)
                /rpt_family="PB1D10"
                /complement(27592..27673)
                /rpt_family="PB1D10"
                /complement(27673..27754)
                /rpt_family="PB1D10"
                /complement(27754..27835)
                /rpt_family="PB1D10"
                /complement(27835..27916)
                /rpt_family="PB1D10"
                /complement(27916..28000)
                /rpt_family="PB1D10"
                /complement(28000..28081)
                /rpt_family="PB1D10"
                /complement(28081..28162)
                /rpt_family="PB1D10"
                /complement(28162..28243)
                /rpt_family="PB1D10"
                /complement(28243..28324)
                /rpt_family="PB1D10"
                /complement(28324..28405)
                /rpt_family="PB1D10"
                /complement(28405..28486)
                /rpt_family="PB1D10"
                /complement(28486..28567)
                /rpt_family="PB1D10"
                /complement(28567..28648)
                /rpt_family="PB1D10"
                /complement(28648..28729)
                /rpt_family="PB1D10"
                /complement(28729..28810)
                /rpt_family="PB1D10"
                /complement(28810..28891)
                /rpt_family="PB1D10"
                /complement(28891..28972)
                /rpt_family="PB1D10"
                /complement(28972..29053)
                /rpt_family="PB1D10"
                /complement(29053..29134)
                /rpt_family="PB1D10"
                /complement(29134..29215)
                /rpt_family="PB1D10"
                /complement(29215..29296)
                /rpt_family="PB1D10"
                /complement(29296..29377)
                /rpt_family="PB1D10"
                /complement(29377..29458)
                /rpt_family="PB1D10"
                /complement(29458..29539)
                /rpt_family="PB1D10"
                /complement(29539..29620)
                /rpt_family="PB1D10"
                /complement(29620..29701)
                /rpt_family="PB1D10"
                /complement(29701..29782)
                /rpt_family="PB1D10"
                /complement(29782..29863)
                /rpt_family="PB1D10"
                /complement(29863..29944)
                /rpt_family="PB1D10"
                /complement(29944..30025)
                /rpt_family="PB1D10"
                /complement(30025..30106)
                /rpt_family="PB1D10"
                /complement(30106..30187)
                /rpt_family="PB1D10"
                /complement(30187..30268)
                /rpt_family="PB1D10"
                /complement(30268..30349)
                /rpt_family="PB1D10"
                /complement(30349..30430)
                /rpt
```

```

repeat_region complement(3157..3276)
repeat_region /rpt_unit=GA
repeat_region complement(3328..3359)
repeat_region /rpt_family="MIR"
repeat_region 3890..3962
repeat_region /rpt_unit=TAAA
repeat_region complement(5643..5705)
repeat_region /rpt_family="PB1D10"
repeat_region 7199..7335
repeat_region /rpt_family="MMB1"
repeat_region complement(7390..7480)
repeat_region /rpt_family="MMB1F"
repeat_region complement(7538..7645)
repeat_region /rpt_family="MMSOR2"
repeat_region complement(7674..8071)
repeat_region /rpt_family="MTA"
repeat_region complement(8071..8121)
repeat_region /rpt_family="PB1D10"
repeat_region complement(8246..8349)
repeat_region /rpt_family="PB1D10"
repeat_region 8717..8756
repeat_region /rpt_family="GC-rich"
repeat_region complement(<8978..>9017)
repeat_region /product="unknown"
repeat_region complement(<8978..>9017)
repeat_region /note="unknown: Similar to ESTs AA260075 and AA144156.
repeat_region Similar to an exon in cDNAs in L26528 and D01046"
repeat_region /codon_start=1
repeat_region /protein_id="AAC97964.1"
repeat_region /db_xref="GI:4050091"
repeat_region /translation="MGTTRDDEYLFK"
repeat_region complement(9622..9702)
repeat_region /rpt_unit=GA
repeat_region complement(9775..9811)
repeat_region /rpt_family="POLY_A"
repeat_region complement(9812..9950)
repeat_region /rpt_family="MMB1F"
repeat_region complement(9965..10006)
repeat_region /rpt_unit=CA
repeat_region 1047..10159
repeat_region /rpt_unit=GA
repeat_region 10434..10612
repeat_region /rpt_family="MMB2"
repeat_region complement(10753..10952)
repeat_region /rpt_family="B3"
repeat_region 11202..11330
repeat_region /rpt_family="B3"
repeat_region complement(11348..11433)
repeat_region /rpt_unit=GA
repeat_region complement(11468..11605)
repeat_region /rpt_family="MMB1"
repeat_region complement(11627..11770)
repeat_region /rpt_family="MMB1F"
repeat_region complement(11909..12232)
repeat_region /rpt_family="ORR1A1"
repeat_region 12413..12472
repeat_region /rpt_family="MMB1F"
repeat_region 13085..13229
repeat_region /rpt_family="MTD"
repeat_region complement(13321..13598)
repeat_region /rpt_family="MMETNLTR"
repeat_region 13597..13778
repeat_region /rpt_family="MTD"
repeat_region complement(13948..14046)
repeat_region /rpt_family="MMB1F"
repeat_region 14293..14381
repeat_region /rpt_family="PB1"
repeat_region 14442..14558
repeat_region /rpt_family="MMB1"
repeat_region 14598..14673
repeat_region /rpt_family="B4"
repeat_region 14629..14724
repeat_region /rpt_family="MMSOR2"

```

```

repeat_region 14851..14991
repeat_region /rpt_family="ORR1B-INT"
repeat_region 14880..15016
repeat_region /rpt_family="MMB1"
repeat_region 15781..15911
repeat_region /rpt_family="MMB1F"
repeat_region complement(15997..16184)
repeat_region /rpt_family="B3"
repeat_region 16299..16422
repeat_region /rpt_family="MMB1F"
repeat_region 16429..16463
repeat_region /rpt_unit=CA
repeat_region 16479..16707
repeat_region /rpt_family="MMB2"
repeat_region 16719..16923
repeat_region /rpt_family="B3"
repeat_region complement(17436..17469)
repeat_region /rpt_family="AT-rich"
repeat_region complement(17472..17515)
repeat_region /rpt_unit=CA
repeat_region complement(17558..17625)
repeat_region /rpt_family="B4"
repeat_region 17707..17824
repeat_region /rpt_family="PB1D9"
repeat_region complement(17892..17976)
repeat_region /rpt_family="MMSOR2"
repeat_region complement(18032..18164)
repeat_region /rpt_unit=GA
repeat_region complement(18165..18233)
repeat_region /rpt_family="MMB2"
repeat_region complement(18229..18274)
repeat_region /rpt_family="B3"
repeat_region complement(18327..18499)
repeat_region /rpt_family="MMB2"
repeat_region 18570..18781
repeat_region /rpt_family="B4"
repeat_region 18656..18781
repeat_region /rpt_family="PB1D10"
repeat_region complement(18791..18827)
repeat_region /rpt_unit=GA
repeat_region complement(18828..18972)
repeat_region /rpt_family="MMB1"
repeat_region 19004..19085
repeat_region /rpt_family="B4A"
repeat_region 19097..19266
repeat_region /rpt_family="MMB2"
repeat_region 19303..19402
repeat_region /rpt_family="B4A"
repeat_region 19402..19551
repeat_region /rpt_family="MMB1"
repeat_region 19556..19594
repeat_region /rpt_unit=TAAA
repeat_region complement(20527..20633)
repeat_region /rpt_family="MMB1"
repeat_region complement(20740..20769)
repeat_region /rpt_unit=GA
repeat_region complement(20810..20831)
repeat_region /rpt_family="AT-rich"
repeat_region complement(20846..20945)
repeat_region /rpt_family="B3"
repeat_region complement(21057..21094)
repeat_region /rpt_family="MUSID5"
repeat_region complement(21115..21239)
repeat_region /rpt_family="PB1D7"
repeat_region 21474..21553
repeat_region /rpt_family="B3"
repeat_region 21570..21656

```

Query Match 72.9%; Score 22.6; DB 10; Length 213245;  
 Best Local Similarity 86.2%; Pred. No. 4.2;  
 Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cctctagaccagatgcattgttgc 30





```

repeat_region /rpt_family="MMB2"
10369..10460 /rpt_family="(GA)n"
unsure 10454..10474 /gene="tapasin"
repeat_region /note="low quality data"
complement(10836..10943)
/rpt_family="PBD17"
repeat_region 10973..11020 /rpt_family="MUSID5"
repeat_region 11215..11328 /rpt_family="MMB1F"
repeat_region 11329..11427 /rpt_family="(GA)n"
complement(11428..11459)
/rpt_family="(CA)n"
repeat_region 11467..11637 /rpt_family="B3"
complement(113691..13830)
repeat_region /rpt_family="MMB1"
complement(13859..13993)
/rpt_family="MMSOR2"
complement(14790..14932)
/rpt_family="MMB1F"
complement(14947..15087)
/rpt_family="MMB1"
complement(15009..15110)
/rpt_family="MTE2"
15133..15243
/rpt_family="MMB1"
complement(15387..15490)
/rpt_family="URR1"
101n(16444..16522,16723..16919,18204..18287,18518..18696,
18799..18849,18932..19229,19487..19738,19829..19932,
20019..20103,20190..20258,20375..20449,21453..21485,
21612..21733,21838..21933,22023..22134,22218..22511,
23341..23455,23543..23754)
/gene="RLF"
/product="RalGDS-like factor"
16444..24873
/gene="RLF"
/note="Described as a Ras-associating protein in GenBank
Accession Number U54639"
join(16764..16919,18204..18287,18518..18696,18799..18849,
18932..19229,19487..19738,19829..19932,20019..20103,
20190..20258,20375..20449,21453..21485,21612..21733,
21838..21933,22023..22134,22218..22511,23341..23455,
23543..23754)
/gene="RLF"
/note="intron-exon boundaries defined in relation to cDNA
found in GenBank Accession Number U54639"
/codon_start=1
/product="RalGDS-like factor"
/db_xref="GI:3811378"
translation="MLPRPLRLLLDTTPGGVVLSSFRSDPEEGDPGAVGGQEE
EDEEEESVSWDEEGATFTVTSROYRLDPLAPLPSPSSRRRLRACGLLEALVR
HLDDATAGADMWFPALLATHRAFTSTPALGLVADRLALEASPPGELERTGVAI
SVLSTWASHPEDGSEVAGQRLRLSFLLRTGYAAREGVGGSADLIRNLRARVDR
APDLPLALPDGSPADPTDVLVFLADHLAEQLTLDLAEFLNLNIPSQGLGLWGHDR
72.9%; Score 22.6; DB 10; Length 273800;
Query Match
Best Local Similarity 86.2%; Pred. No. 4.2;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 cctctagagccagagtcattgttgc 30
||||| |||||||
Db 4020 CCTCTACCTCCAGGATGATTGTGCC 4048

```

```

RESULT 11
163542/c 163542
LOCUS 59 bp DNA linear PAT 07-OCT-1997

```

```

DEFINITION Sequence 5 from patent US 5663141.
ACCESSION I63542
VERSION I63542.1 GI:2481115
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 59)
AUTHORS Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
Hoeffken,H.Wolfgang.
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL Patent: US 5663141-A 5 02-SEP-1997;
FEATURES
Location/Qualifiers
source 1..59
/organism="unknown"
BASE COUNT 19 a 14 c 16 g 10 t
ORIGIN

```

```

Query Match 71.6%; Score 22.2; DB 6; Length 59;
Best Local Similarity 88.9%; Pred. No. 5.5;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 gctctagagccagagtcattgtt 27
||||| |||||||
Db 27 GCCGTCAGAGCCAGGATGCAATTGTT 1

```

```

RESULT 12
LOCUS A18858 90 bp DNA linear PAT 20-APR-1994
DEFINITION Oligonucleotide 2 HVL.
ACCESSION A18858
VERSION A18858.1 GI:513445
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 90)
TITLE METHOD FOR THE RECOMBINANT PRODUCTION OF HIRIDIUS AND HIRUDIN-LIKE
POLYPEPTIDES
JOURNAL Patent: WO 91/17250-A 2 14-NOV-1991;
FEATURES
Location/Qualifiers
source 1..90
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 22 a 27 c 19 g 22 t
ORIGIN

```

```

Query Match 71.0%; Score 22; DB 6; Length 90;
Best Local Similarity 83.3%; Pred. No. 7;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 2 cctctagagccagagtcattgttgc 31
||||| |||||||
Db 15 CCGTCAGAACCCAGGATGCAATTGTACCC 44

```

```

RESULT 13
LOCUS A18862 90 bp DNA linear PAT 20-APR-1994
DEFINITION Oligonucleotide 2 HV12.
ACCESSION A18862
VERSION A18862.1 GI:513449
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 90)
TITLE METHOD FOR THE RECOMBINANT PRODUCTION OF HIRIDIUS AND HIRUDIN-LIKE

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:53 ; Search time 758.37 Seconds  
(without alignments)  
70.183 Million cell updates/sec

Title: US-10-053-641-6  
Perfect score: 31  
Sequence: 1 gctctagaccagatgcattgttgc 31

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 347872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	29.4	94.8	468	AAQ93244	Fusion construct o
C 2	26.2	84.5	726	AAQ12379	Hirudin peptide/Pr
C 3	22	71.0	183	AAQ14927	Partial Hirudin HV
C 4	22	71.0	185	AAQ14926	Synthetic hirudin
C 5	22	71.0	196	AAQ70319	Sequence encoding
C 6	22	71.0	217	AAQ60355	Desulphatohirudin
C 7	22	71.0	217	AAQ70323	Sequence of the de
C 8	22	71.0	224	AAQ60746	DNA encoding hirud
C 9	22	71.0	279	AAQ13831	MSP signal peptide

C 10	20.8	67.1	201	12	AAQ12153	Synthetic hirudin
C 11	20.8	67.1	201	13	AAQ25184	Hirudin HV-1. Syn
C 12	20.8	67.1	420	12	AAQ12155	Factor Xa-cleavabl
C 13	20.8	67.1	1458	12	AAQ12162	Factor Xa-cleavabl
C 14	20.8	67.1	1467	12	AAQ12490	Factor Xa-cleavabl
C 15	20.4	65.8	2300	24	AAQ62246	CDNA sequence #33
C 16	20.4	65.8	2864	22	AAH76458	CDNA corresponding
C 17	20.2	65.2	1618	21	AAQ44088	Zea mays DNA fragm
C 18	20	64.5	9048	21	AAA64507	Nucleotide sequenc
C 19	20	64.5	10556	22	AAI59459	Human polyuucleoti
C 20	19.8	63.9	828	13	AAQ24269	Encodes PLGF angio
C 21	19.8	63.9	3543	22	AAH18248	Human CDNA sequenc
C 22	19.6	63.2	104	21	AAQ13076	Human secreted pro
C 23	19.2	61.9	227	22	AAH61507	S. marcescens hiru
C 24	19.2	61.9	229	9	AAH81294	Sequence encoding
C 25	19.2	61.9	238	9	AAH81295	Sequence encoding
C 26	19.2	61.9	1458	22	AAH68269	C glutamicum codin
C 27	19.2	61.9	309400	22	AAH68534	C glutamicum codin
C 28	19	61.3	969	23	ABL25309	Drosophila melanog
C 29	19	61.3	1817	22	AAQ21961	Human collagen gen
C 30	19	61.3	2906	23	ABL25306	Drosophila melanog
C 31	19	61.3	2969	23	ABL25308	Drosophila melanog
C 32	19	61.3	10234	22	AAQ105428	Human reproductive
C 33	19	61.3	20084	22	AAQ21773	Human gene for col
C 34	19	61.3	20084	22	AAH83656	Nucleotide sequenc
C 35	19	61.3	35959	22	AAK78275	Human immune/haema
C 36	19	61.3	50885	22	AAK70336	Human immune/haema
C 37	18.8	60.6	198	13	AAQ24986	Sequence encoding
C 38	18.8	60.6	210	12	AAQ10452	HV-1 gene. Synthe
C 39	18.8	60.6	430	22	AAQ25702	Human breast cance
C 40	18.8	60.6	453	22	AAQ18097	Human breast cance
C 41	18.8	60.6	456	22	AAQ17956	Human breast cance
C 42	18.8	60.6	552	23	ABL10237	Drosophila melanog
C 43	18.8	60.6	665	21	AAQ43982	Mouse secreted exp
C 44	18.8	60.6	984	19	AAV20815	Rat haematopoietic
C 45	18.8	60.6	1360	24	AB199523	Mouse ischaemic co

ALIGNMENTS

RESULT 1  
AAQ93244/c  
ID AAQ93244 standard; DNA; 468 BP.

AC AAQ93244;

XX 01-NOV-1995 (first entry)

DT Fusion construct of glucoamylase-hirudin DNA.

DE expression cassette; recombinant protein; production; Hansenula;  
KW yeast; leader sequence; adaptor; alpha helix; glucoamylase;  
KW secretion; processing; thrombin inhibitor; hirudin; ds.

OS Synthetic.

XX Key

FT primer\_bind

FT Location/Qualifiers

FT complement (1..33)

FT /tag= a

FT /note= "primer AAQ85840 binding site"

FT CDS

FT 13..444

FT /tag= b

FT /product= glucoamylase-hirudin fusion protein

FT 199..223

FT /tag= c

FT /note= "primer AAQ85841 binding site"

FT complement (220..267)

FT /tag= d

FT /note= "primer AAQ85842 binding site"

FT 13..228

FT /tag= e

FT /note= "encodes amino acids 1-72 of glucoamylase"

FT misc\_feature 241..246  
 FT /\*tag= f  
 FT /note= "encodes processor signal"  
 FT misc\_feature 247..444  
 FT /\*tag= g  
 FT /note= "encodes hirudin-h120"  
 FT primer\_bind 0..468  
 FT /\*tag= h  
 FT /note= "primer AAQ85843 binding site"  
 XX  
 PN DE4329969-A.  
 XX  
 XX  
 PD 09-MAR-1995.  
 XX  
 XX 04-SEP-1993; 93DE-4329969.  
 XX  
 XX 04-SEP-1993; 93DE-4329969.  
 XX  
 PA (BADI ) BASF AG.  
 PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.  
 XX  
 PI Bollschweiler C, Janowicz ZA, Piontek M, Schweden J;  
 PI Strasserawm, Weydemann U;  
 XX  
 XX WPI; 1995-107810/15.  
 DR P-PSDB; AAR76951.  
 XX

PT Recombinant protein prodn. in Hansenula yeast - transformed with  
 PT expression cassette contg. leader, adaptor, processing signal and  
 PT gene, provides efficient secretion and correct processing  
 XX  
 PS Example 1; Fig 1; 10pp; German.  
 XX  
 CC AAQ93244 is a DNA fusion product resulting from ligation of an  
 CC EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding  
 CC a leader sequence, an adaptor (see AAR71472, amino acids 23-72 of GAM  
 CC (glucanylase from Schwannomyces occidentalis, plus -His-Pro-Leu-cln at  
 CC the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,  
 CC encoding hirudin. AAQ93244 is an example of an expression vector insert  
 CC of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;  
 CC P = processor; Gene = structural gene). The cassettes ensure efficient  
 CC secretion and correct processing of heterologous structural genes in  
 CC yeast of the genus Hansenula, and so provides high yields of mature  
 CC proteins and facilitates subsequent purification.  
 XX  
 SQ Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match 94.8%; Score 29.4; DB 16; Length 468;  
 Best Local Similarity 96.8%; Pred. No. 0.0015;  
 Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcctctagagccagcagcattgtgtgcc 31  
 ||| |||||  
 Db 348 GCCTTTAGAGCCAGGATGCATTGTGTGCC 318

RESULT 2  
 AAQ12379/C  
 ID AAQ12379 standard; DNA; 726 BP.  
 XX  
 AC AAQ12379;  
 XX  
 XX 17-SEP-1991 (first entry)  
 XX  
 DE Hirudin peptide/Protein A fusion gene.  
 XX  
 KW anticoagulant; fusion protein; ds.  
 XX  
 XX Key Location/Qualifiers  
 FT mat\_peptide 1..489  
 FT /\*tag= a  
 FT /product= Protein A

FT mat\_peptide 529..720  
 FT /\*tag= b  
 FT /product= Hirudin  
 FT misc\_RNA 490..528  
 FT /\*tag= c  
 FT /product= linker oligopeptide Y

XX DE3942580-A.  
 PN 27-JUN-1991.  
 XX  
 PD 22-DEC-1989; 89DE-3942580.  
 XX  
 PF 22-DEC-1989; 89DE-3942580.  
 XX  
 PR (BADI ) BASF AG.  
 XX  
 XX Korwer W;  
 XX  
 XX WPI; 1991-1942336/27.  
 DR P-PSDB; AAR12751.  
 XX

XX Hirudin peptide prodn. by cleaving new fusion peptide - of  
 PT hirudin and protein A, expressed in high yield and stable,  
 PT soluble form by transformed E.coli  
 PT  
 XX Example 1; Page 6-7; 9pp; German.

XX This sequence is an example of a fusion construct for expression of  
 CC the fusion peptide of the invention. The Met residue in the linker  
 CC oligopeptide allows cleavage by CNBr to release two fragments which  
 CC can be easily separated by IgG affinity chromatography. (The protein  
 CC A component binds to IgG sepharose). Increased yields of hirudin are  
 CC obtained using this fusion construct.  
 XX  
 SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match 84.5%; Score 26.2; DB 12; Length 726;  
 Best Local Similarity 90.3%; Pred. No. 0.045;  
 Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 gcctctagagccagcagcattgtgtgcc 31  
 ||| |||||  
 Db 627 GCCTCAGAGCCAGGATGCATTGTGTGCC 597

RESULT 3  
 AAQ14927/C  
 ID AAQ14927 standard; DNA; 183 BP.  
 XX  
 AC AAQ14927;  
 XX  
 XX 24-FEB-1992 (first entry)  
 DT  
 DE Partial Hirudin HV12 coding sequence.  
 XX  
 XX anti-thrombin activity; thrombosis; HV1; HV2; blood clotting; ds.  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FT CDS 7..177  
 FT /\*tag= a  
 FT /note= "encodes residues 10-65 of HV12"  
 XX  
 PN W09117250-A.  
 XX  
 PD 14-NOV-1991.  
 XX  
 XX 05-APR-1991; 91WO-EP00643.  
 XX  
 XX 10-MAY-1990; 90GB-0010552.



Db 101 CCGTCAGAACCCAGGATGTCATTGTTACCC 72

## RESULT 6

AAN60355/c  
ID AAN60355 standard; DNA; 217 BP.

AC AAN60355;

DT 20-JUN-1991 (first entry)

DE Desulphatohirudin (II).

KW Desulphatohirudin; antibodies; thrombin; ds.

OS Synthetic.

FH Key Location/Qualifiers

CDS 10..207

FT /\*tag= a

TT /product= hirudin

EP168342-A.

PD 15-JAN-1986.

PF 10-JUN-1985; 85EP-0810268.

PR 14-JUN-1984; 84CH-0288284.

XX (CIBA ) CIBA GEIGY AG.

XX Liersch M, Rink H, Marki W, Grutter MG, Meyhack B;

XX WPI; 1986-015589/03.

DR P-PSDB; AAP60395.

XX DNA sequences coding for hirudin and derivs. - and expression  
PT vectors, transformed cells, monoclonal antibodies and hybridomas,  
PT useful as thrombin inhibitor.

PS Disclosure; Page 10; 123pp; German.

XX The sequence comprises an EcoRI restriction enzyme site at the  
CC 5'-end and a BamHI restriction enzyme site at the 3'-end.  
CC The sequence may be introduced into a vector for the transformation  
CC of hosts, e.g. E. coli. The hirudin can thus be prepared on a  
CC large scale. The product and its derivs. are thrombin inhibitors,  
CC useful in anticoagulant therapy, esp. when injected at doses  
CC of 0.01-0.05 mg/kg.

XX Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match 71.0%; Score 22; DB 7; Length 217;

Best Local Similarity 83.3%; Pred. No. 2.7;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccagagtcattgttgc 31

Db 113 CCGTCAGAACCCAGGATGTCATTGTTACCC 84

## RESULT 7

AAN70323/c

ID AAN70323 standard; DNA; 217 BP.

XX AAN70323;

DT 02-APR-1991 (first entry)

DE Sequence of the desulphatohirudin gene.

XX

KW Anticoagulant; thrombin inhibitor; ds.

XX EP225633-A.

PD 16-JUN-1987.

XX 09-DEC-1986; 86EP-0117098.

XX 29-MAY-1986; 86GB-0013088.

PR 12-DEC-1985; 85GB-0030631.

XX (CIBA ) CIBA GEIGY AG.

PA (PLAN-) PLANTORGAN WERK HEINRICH.

PA (CHRI-) PLANTORGANW CHRISTENSEN.

XX Meyhack B, Marki W, Helm J;

XX WPI; 1987-164868/24.

XX New DNA constructs and hybrid vectors for transformation of yeast  
PT etc. - useful for prodn. and secretion of protein with hirudin  
PT activity for use as thrombin inhibitors.

PS Example; p44; 146pp; English.

XX The preferred DNA construct of the invention contains the PHO5  
CC promoter and a DNA segment consisting of the PHO5 signal sequence  
CC upstream of and in reading frame with a DNA sequence coding for  
CC mature desulphatohirudin. The segment is under the transcriptional  
CC control of the PHO5 promoter and the 3' flanking sequence of the  
CC PHO5 gene.

XX Sequence 217 BP; 58 A; 53 C; 58 G; 48 T; 0 other;

Query Match 71.0%; Score 22; DB 8; Length 217;

Best Local Similarity 83.3%; Pred. No. 2.7;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccagagtcattgttgc 31

Db 113 CCGTCAGAACCCAGGATGTCATTGTTACCC 84

## RESULT 8

AAN60746/c

ID AAN60746 standard; DNA; 224 BP.

XX AAN60746;

XX 01-JAN-1980 (first entry)

DT DNA encoding hirudin-like polypeptide.

DE Hirudin; thrombin-antagonist; anticoagulant; ds.

XX Synthetic.

XX DE3445517-A.

XX 19-JUN-1986.

XX 13-DEC-1984; 84DE-3445517.

XX 13-DEC-1984; 84DE-3445517.

XX (GENB-) GEN-BIO-TEC GES GEN.

XX Fortkamp E, Rieger M, Sommer R;

XX WPI; 1986-162802/26.

DR P-PSDB; AAP60827.

XX





Query Match 67.1%; Score 20.8; DB 13; Length 201;  
Best Local Similarity 91.7%; Pred. No. 9.3;  
Matches 22; Conservative 0; Mismatches 2; Indels 0

Db 95 GAACCCAGGATGCATTGTACCC 72

RESOLI 12  
AAQ12155/C

17-SEP-1991 (first entry)  
Factor Xa-cleavable hirudin-IEGR-hirudin gene.

AA	
AA	Fusion protein; blood clotting; coagulation; fibrinolysis;
KW	antithrombotic; thrombolysis; streptokinase; ss.
KW	

XX Synthetic.

XX  
XX

```

FT      key      location/Qualifiers
FT      CDS      1..195
FT      /tag= a

```

	FT	FT	CDS
1997	1.00	1.00	1.00
1998	1.00	1.00	1.00
1999	1.00	1.00	1.00
2000	1.00	1.00	1.00
2001	1.00	1.00	1.00
2002	1.00	1.00	1.00
2003	1.00	1.00	1.00
2004	1.00	1.00	1.00
2005	1.00	1.00	1.00
2006	1.00	1.00	1.00
2007	1.00	1.00	1.00
2008	1.00	1.00	1.00
2009	1.00	1.00	1.00
2010	1.00	1.00	1.00
2011	1.00	1.00	1.00
2012	1.00	1.00	1.00
2013	1.00	1.00	1.00
2014	1.00	1.00	1.00
2015	1.00	1.00	1.00
2016	1.00	1.00	1.00
2017	1.00	1.00	1.00
2018	1.00	1.00	1.00
2019	1.00	1.00	1.00
2020	1.00	1.00	1.00
2021	1.00	1.00	1.00
2022	1.00	1.00	1.00
2023	1.00	1.00	1.00
2024	1.00	1.00	1.00
2025	1.00	1.00	1.00
2026	1.00	1.00	1.00
2027	1.00	1.00	1.00
2028	1.00	1.00	1.00
2029	1.00	1.00	1.00
2030	1.00	1.00	1.00
2031	1.00	1.00	1.00
2032	1.00	1.00	1.00
2033	1.00	1.00	1.00
2034	1.00	1.00	1.00
2035	1.00	1.00	1.00
2036	1.00	1.00	1.00
2037	1.00	1.00	1.00
2038	1.00	1.00	1.00
2039	1.00	1.00	1.00
2040	1.00	1.00	1.00
2041	1.00	1.00	1.00
2042	1.00	1.00	1.00
2043	1.00	1.00	1.00
2044	1.00	1.00	1.00
2045	1.00	1.00	1.00
2046	1.00	1.00	1.00
2047	1.00	1.00	1.00
2048	1.00	1.00	1.00
2049	1.00	1.00	1.00
2050	1.00	1.00	1.00
2051	1.00	1.00	1.00
2052	1.00	1.00	1.00
2053	1.00	1.00	1.00
2054	1.00	1.00	1.00
2055	1.00	1.00	1.00
2056	1.00	1.00	1.00
2057	1.00	1.00	1.00
2058	1.00	1.00	1.00
2059	1.00	1.00	1.00
2060	1.00	1.00	1.00
2061	1.00	1.00	1.00
2062	1.00	1.00	1.00
2063	1.00	1.00	1.00
2064	1.00	1.00	1.00
2065	1.00	1.00	1.00
2066	1.00	1.00	1.00
2067	1.00	1.00	1.00
2068	1.00	1.00	1.00
2069	1.00	1.00	1.00
2070	1.00	1.00	1.00
2071	1.00	1.00	1.00
2072	1.00	1.00	1.00
2073	1.00	1.00	1.00
2074	1.00	1.00	1.00
2075	1.00	1.00	1.00
2076	1.00	1.00	1.00
2077	1.00	1.00	1.00
2078	1.00	1.00	1.00
2079	1.00	1.00	1.00
2080	1.00	1.00	1.00
2081	1.00	1.00	1.00
2082	1.00	1.00	1.00

```
FT /tag= D
FT /label= linker
FT /note= "encodes cleavage site TEGR"
```

```

F1 tag=
F1 label= hirudin gene

```

XX  
PN  
W09109125-AXX  
PD  
27-JUN-1991XX  
PF  
07-DEC-1990. 90W0-GB01911XX  
PR 07-DEC-1990: 90WO-GB01911.

PR 07-DEC-1989; 89GB-002172Z  
XX

XX  
PA (BRBL-) BRIT BIO-TECHN LTD.

PI Dawson KM, Hunter MG, Czaplewski LG;

XX WPI; 1991-208151/28.  
DR P-PSDB; AAR12888.  
XX  
PT Fusion protein cleavage by blood clotting enzyme - for prodn. of  
PT fractions having greater antithrombotic activity for therapy and  
PT prophylaxis.

PT prophylaxis.

Disclosure; Page 78; 115pp; English.

The sequence of the synthetic hirudin HV-1 genes was designed based on the published amino acid sequence (Dodd J., et al FEBS Letters 165 180 (1984)). Unique restriction sites were incorporated to facilitate subsequent genetic manipulation. The codons selected were those favourable for *S. cerevisiae* or *E. coli*. The sequence was divided into 12 oligomers which were synthesised and then annealed. The ligation prod. was ligated to HindIII and EcoRI treated pUC19 plasmid DNA and the resulting vector used to transform *E. coli* K12 HBW7. Plasmid pUC19 HV-1 was isolated from transformants and inserted into plasmid pSW6, a shuttle vector, for expression. The gene was then used to construct an expression vector in which two hirudin genes are linked together via a linker encoding a cleavage site for factor Xa. The factor Xa is present at the site of the target thrombus so the active agents are released specifically at the place where clot formation is occurring.

in which two fibrinogen genes are linked together via a linker encoding a cleavage site for factor Xa. The factor Xa is present at the site of the target thrombus so the active agents are released specifically at the place where clot formation is occurring.

The sequence of the synthetic tirudin HV-1 genes was designed based on the published amino acid sequence (Dodt, J. et al. FEBS Letters 165 180 (1984)). The sequence of streptokinase was obtained from PCR amplified chromosomal DNA from *S. equimillilis* ATCC 10009, or ATCC 9542. The primers used for the PCR were based on the published DNA sequence of *S. equimillilis* strain H46A (Malke, H., Roe, B., and Ferretti, J. J. Gene 34 357-362 (1985)). The two sequences were used to construct an expression vector in which the

CC or ATCC 9642. The primers used for the PCR were based on the pub-  
 CC lished DNA sequence of *S. equismilis* strain H46A (Walke, H., Roe,  
 CC B., and Ferretti, J.J., Gene 34: 357-362 [1985]). The two  
 CC sequences were used to construct an expression vector in which the  
 CC streptokinase gene is linked to the hirudin gene via a linking  
 CC sequence encoding a cleavage site for factor Xa. The factor Xa is  
 CC present at the site of the target thrombus so the active agents are  
 CC released specifically at the place where clot formation is occurring.  
 CC See also AAQ12153-Q12156 and AAQ12158-Q12162.  
 XX  
 SQ Sequence 1467 BP; 494 A; 317 C; 292 G; 364 T; 0 other;

Query Match 67.1%; Score 20.8; DB 12; Length 1467;  
 Best Local Similarity 91.7%; Pred. No. 13;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 gagccaggatgcatgtgtgccc 31  
 || |||||  
 Db 1349 GAACCCAGGATGCAATTGTTACCC 1326

RESULT 15  
 AAS62246  
 ID AAS62246 standard; cDNA; 2300 BP.

XX AC AAS62246;

DT 14-FEB-2002 (first entry)

DE cDNA sequence #33 encoding novel human secreted protein.

KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
 KW immune deficiency disorder; blood disorder; inflammatory disorder;  
 KW infectious disorder; gene therapy; antimicrobial; hepatotropic;  
 KW immunosuppressive; antirheumatic; ss.

XX OS Homo sapiens.

XX PN WO200177291-A2.

XX PD 18-OCT-2001.

XX PF 29-MAR-2001; 2001WO-US10485.

XX PR 06-APR-2000; 2000US-195604P.

XX PA (GEMY ) GENETICS INST INC.

XX PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
 PI Gulukota K, Graham JR;

XX DR WPI; 2002-010900/01.

XX PT New polynucleotides encoding secreted proteins useful for treating e.g.  
 XX asthma, HIV and Crohn's disease -

XX PS Claim 1; Page 88-89; 391pp; English.

XX The present invention relates to the isolation of novel cDNA sequences  
 CC which encode human secreted proteins. The cDNA sequences have been  
 CC derived from a variety of human tissues. The invention also provides  
 CC a method for producing proteins from these polynucleotide sequences.  
 CC The proteins are useful for identifying compounds that modulate their  
 CC activity and production, and the cell is also useful for identifying  
 CC compounds that modulate expression of the polynucleotide sequences  
 CC encoding the secreted proteins. The sequences of the invention are  
 CC useful for treating diseases such as hyperproliferative disorders  
 CC (e.g. cancer), immune deficiency disorders (e.g. severe combined  
 CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple  
 CC sclerosis), blood disorders (e.g. thrombocytopenia), inflammatory  
 CC disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).  
 CC The polynucleotide sequences of the invention are also useful in gene

CC therapy. AAS62214-AAS62838 represent the cDNA sequences of the  
 CC invention that encode for novel human secreted proteins.  
 XX  
 SQ Sequence 2300 BP; 619 A; 545 C; 607 G; 529 T; 0 other;

Query Match 65.8%; Score 20.4; DB 24; Length 2300;  
 Best Local Similarity 80.0%; Pred. No. 22;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcctctagagcccgagatgcattgtgtgcc 30  
 | ||| ||| ||||| ||| |||||  
 Db 448 ggctccagaacccaggagaagaattgtgcc 477

Search completed: July 15, 2002, 22:44:55  
 Job time: 10068 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:31:51 ; Search time 174.99 Seconds<sup>3</sup>  
(without alignments)  
43.515 Million cell updates/sec

Title: US-10-053-641-6

Sequence: 1 gcctctagagccaggatgcatgtgtgccc 31

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/1na/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/1na/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/1na/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/1na/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/1na/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/1na/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	26.2	84.5	235	1	US-08-262-384A-7
C 2	26.2	84.5	235	1	US-08-262-384A-8
C 3	26.2	84.5	726	1	US-08-262-384A-9
C 4	26.2	84.5	726	1	US-08-262-384A-10
C 5	25.4	81.9	59	1	US-07-859-453E-9
C 6	25.4	81.9	59	1	US-07-859-453E-11
C 7	25.4	81.9	129	1	US-08-262-384A-3
C 8	22.2	71.6	59	1	US-07-859-453E-5
C 9	22.2	71.6	59	1	US-07-859-453E-13
C 10	22.2	71.6	217	6	5422249-13
C 11	22.2	71.6	217	6	5422249-4
C 12	21.8	70.3	279	1	US-08-186-222-3
C 13	21.8	70.3	64	6	5422249-11
C 14	21.8	70.3	109	6	5422249-6
C 15	20.8	67.1	40	1	US-07-854-596B-54
C 16	20.8	67.1	40	1	US-07-854-596B-55
C 17	20.8	67.1	201	1	US-07-854-596B-1
C 18	20.8	67.1	223	1	US-07-854-596B-7
C 19	20.8	67.1	420	1	US-07-854-596B-8
C 20	20.8	67.1	1458	1	US-07-854-596B-42
C 21	19.8	63.9	1467	1	US-07-854-596B-46
C 22	19.8	63.9	29	2	US-08-560-098A-39
C 23	19.2	61.9	826	2	US-08-039-297B-3
C 24	19.2	61.9	128	1	US-08-262-384A-5
C 25	19.2	61.9	212	6	5180668-10
C 26	19.2	61.9	227	2	US-07-982-064-8
C 27	19.2	61.9	8491	2	US-08-757-439-1
C 28	19.2	61.9	59	1	US-07-859-453E-7

28 19 61.3 1817 4 US-08-943-731-193  
29 19 61.3 20084 4 US-08-943-731-5  
c 30 18.8 60.6 198 1 US-07-910-528-2  
c 31 18.8 60.6 198 1 US-08-348-972-2  
c 32 18.8 60.6 227 6 5164304-3  
c 33 18.8 60.6 228 6 5179196-3  
c 34 18.8 60.6 2115 2 US-08-767-026-3  
c 35 18.6 60.0 198 2 US-08-861-459-2  
c 36 18.4 59.4 195 1 US-08-044-506B-27  
c 37 18.4 59.4 1276 1 US-08-417-460-7  
c 38 18.2 58.7 695 1 US-08-592-126-80  
c 39 18.2 58.7 1269 1 US-08-808-641-2  
c 40 18.2 58.7 1269 2 US-09-064-839-2  
c 41 18.2 58.7 1289 3 US-09-351-438-2  
c 42 17.8 57.4 195 1 US-08-406-948A-7  
c 43 17.8 57.4 204 2 US-08-492-343-1  
c 44 17.8 57.4 204 3 US-09-080-865-2  
c 45 17.8 57.4 204 5 PCT-US94-10048-1

## ALIGNMENTS

RESULT 1  
US-08-262-384A-7/c  
: Sequence 7, Application US/08262384A  
: Patent No. 5624822  
: GENERAL INFORMATION:  
: APPLICANT: Koerwer, Wolfgang  
: TITLE OF INVENTION: The Preparation of Hirusin  
: NUMBER OF SEQUENCES: 10  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Keil & Weinkauff  
: STREET: 1101 Connecticut Avenue  
: CITY: Washington  
: STATE: D.C.  
: COUNTRY: USA  
: ZIP: 20036  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
: OPERATING SYSTEM: IBM AT-compatible, 80486 processor  
: SOFTWARE: WordPerfect version 5.1  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/262,384A  
: FILING DATE: 20-JUN-1994  
: CLASSIFICATION: 530  
: CLASSIFICATION: C 12 N 15/62  
: CLASSIFICATION: C 12 N 15/31  
: CLASSIFICATION: C 07 K 7/10  
: CLASSIFICATION: C 12 P 21/02  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/EP90/02084  
: FILING DATE: 04-DEC-1990  
: APPLICATION NUMBER: US 07861820  
: FILING DATE: 18-JUN-1992  
: INFORMATION FOR SEQ ID NO: 7:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 235 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: US-08-262-384A-7

Query Match 84.5%; Score 26.2; DB 1; Length 235;  
Best Local Similarity 90.3%; Pred. No. 0.004;  
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 gcctctagagccaggatgcatgtgtgccc 31  
Db 135 GCGCTCAGAGCCCGAGGATGCTTTGTTGCC 105

```

RESULT 2
US-08-262-384A-8
; Sequence 8, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-262-384A-8

```

```

Query Match      84.5%; Score 26.2; DB 1; Length 235;
Best Local Similarity 90.3%; Pred. No. 0.004;
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 gctctagagccaggatgcattgttgc 31
   ||| ||||| ||||| ||||| |||||
Db 105 GCCGTCAGAGCCAGGATGCGATTTGTC 135

```

```

RESULT 3
US-08-262-384A-9/c
; Sequence 9, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-262-384A-9

```

```

Query Match      84.5%; Score 26.2; DB 1; Length 726;
Best Local Similarity 90.3%; Pred. No. 0.005;
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 gctctagagccaggatgcattgttgc 31
   ||| ||||| ||||| ||||| |||||
Db 627 GCCGTCAGAGCCAGGATGCGATTTGTC 597

```

```

RESULT 4
US-08-262-384A-10
; Sequence 10, Application US/08262384A
; Patent No. 5624822
; GENERAL INFORMATION:
; APPLICANT: Koerwer, Wolfgang
; TITLE OF INVENTION: The Preparation of Hirudin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-262-384A-10

```

```

Query Match      84.5%; Score 26.2; DB 1; Length 726;

```

Best Local Similarity 90.3%; Pred. No. 0.005;  
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gctcttagagccaggatgcattgtgtgcc 31  
Db 100 GCCCTCAGACGCCAGGATGATGTTGGCC 130

RESULT 5  
US-07-859-453E-9/c  
; Sequence 9, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schweden, Juergen  
; APPLICANT: Hoeffken, Hans Wolfgang  
; TITLE OF INVENTION: Hirudin/polyalkylene glycol  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauf  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS version 7.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/859,453E  
; FILING DATE: 29-MAY-1992  
; CLASSIFICATION: A 61 K 37/64  
; CLASSIFICATION: A 61 K 47/48  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/01998  
; FILING DATE: 22-NOV-1990  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 59 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
US-07-859-453E-9

Query Match 81.9%; Score 25.4; DB 1; Length 59;  
Best Local Similarity 96.3%; Pred. No. 0.0071;  
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gctcttagagccaggatgcattgtgt 27  
Db 27 GCCTTTAGAGCCAGGATGATTTGTT 1

RESULT 6  
US-07-859-453E-11/c  
; Sequence 11, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schweden, Juergen  
; APPLICANT: Hoeffken, Hans Wolfgang  
; TITLE OF INVENTION: Hirudin/polyalkylene glycol

; TITLE OF INVENTION: conjugates  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauf  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS version 7.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/859,453E  
; FILING DATE: 29-MAY-1992  
; CLASSIFICATION: 514  
; CLASSIFICATION: A 61 K 37/64  
; CLASSIFICATION: A 61 K 47/48  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/01998  
; FILING DATE: 22-NOV-1990  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 59 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
US-07-859-453E-11

Query Match 81.9%; Score 25.4; DB 1; Length 59;  
Best Local Similarity 96.3%; Pred. No. 0.0071;  
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gctcttagagccaggatgcattgttt 27  
Db 27 GCCTTTAGAGCCAGGATGATTTGTT 1

RESULT 7  
US-08-262-384A-3/c  
; Sequence 3, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauf  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820





```

; APPLICANT: Schmitz, Albert
; TITLE OF INVENTION: Bacterial Vectors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186,222
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,205
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GB 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, JoAnn
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (recombinant)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Hybrid gene: MSP signal (base pair 1 to 81)/
; INDIVIDUAL ISOLATE: desulfatohirudin (base pair 82 to 279)
; IMMEDIATE SOURCE:
; CLONE: pUCRS/pML310
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..276
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 82..276
; US-08-186-222-3

Query Match 71.0%; Score 22; DB 1; Length 279;
Best Local Similarity 83.3%; Pred. No. 0.33;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccagatgcatgtgtgccc 31
||| ||| ||| ||| ||| ||| ||| |||
Db 182 CCGTCAGAACCCAGGATGTCATTTGTACCC 153

RESULT 12
5422249-11
; Patent No. 5422249
; APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,
; MARKUS G.; MEYHACK, BERND
; TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN
; INHIBITORS
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,974
; FILING DATE: 15-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 700,9978
; FILING DATE: 10-MAY-1991
; APPLICATION NUMBER: 582,816
; FILING DATE: 13-SEP-1990
; APPLICATION NUMBER: 211,065
; FILING DATE: 20-JUN-1988
; APPLICATION NUMBER: 744,453
; FILING DATE: 13-JUN-1985
; SEQ ID NO: 6:
; LENGTH: 109
; 5422249-6

Query Match 70.3%; Score 21.8; DB 6; Length 109;
Best Local Similarity 92.0%; Pred. No. 0.34;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 agagccagagatgcatgtgtgccc 31
||| ||| ||| ||| ||| ||| ||| |||
Db 108 AGAACCCAGGATGTCATTTGTACCC 84

RESULT 14
US-07-854-596B-54/C
; Sequence 54, Application US/07854596B
; Patent No. 5434073
; GENERAL INFORMATION:
; APPLICANT: Dawson, Keith M
; APPLICANT: Hunter, Michael G
; APPLICANT: Craplewski, Lloyd G
; TITLE OF INVENTION: Proteins and nucleic acids
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. John J. McDonnell
; STREET: Ten South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606

```

```

; APPLICATION NUMBER: 700,9978
; FILING DATE: 10-MAY-1991
; APPLICATION NUMBER: 582,816
; FILING DATE: 13-SEP-1990
; APPLICATION NUMBER: 211,065
; FILING DATE: 20-JUN-1988
; APPLICATION NUMBER: 744,453
; FILING DATE: 13-JUN-1985
; SEQ ID NO: 11:
; LENGTH: 64
; 5422249-11

Query Match 70.3%; Score 21.8; DB 6; Length 64;
Best Local Similarity 92.0%; Pred. No. 0.31;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 agagccagagatgcatgtgtgccc 31
||| ||| ||| ||| ||| ||| ||| |||
Db 2 agaaccagagatgcatgtgttacc 26

RESULT 13
5422249-6/C
; Patent No. 5422249
; APPLICANT: LIERSCH, MANFRED; RINK, HANS; MARKI, WALTER; GRUTTER,
; MARKUS G.; MEYHACK, BERND
; TITLE OF INVENTION: PROCESS FOR THE MANUFACTURE OF THROMBIN
; INHIBITORS
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,974
; FILING DATE: 15-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 700,9978
; FILING DATE: 10-MAY-1991
; APPLICATION NUMBER: 582,816
; FILING DATE: 13-SEP-1990
; APPLICATION NUMBER: 211,065
; FILING DATE: 20-JUN-1988
; APPLICATION NUMBER: 744,453
; FILING DATE: 13-JUN-1985
; SEQ ID NO: 6:
; LENGTH: 109
; 5422249-6

Query Match 70.3%; Score 21.8; DB 6; Length 109;
Best Local Similarity 92.0%; Pred. No. 0.34;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 agagccagagatgcatgtgtgccc 31
||| ||| ||| ||| ||| ||| ||| |||
Db 108 AGAACCCAGGATGTCATTTGTACCC 84

RESULT 14
US-07-854-596B-54/C
; Sequence 54, Application US/07854596B
; Patent No. 5434073
; GENERAL INFORMATION:
; APPLICANT: Dawson, Keith M
; APPLICANT: Hunter, Michael G
; APPLICANT: Craplewski, Lloyd G
; TITLE OF INVENTION: Proteins and nucleic acids
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. John J. McDonnell
; STREET: Ten South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:49:08 ; Search time 6165.88 Seconds  
(without alignments)  
67.858 Million cell updates/sec

Title: US-10-053-641-6

Perfect score: 31

Sequence: 1 gcctctagagccaggatgattgtgtgcc 31

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hcc.\*  
9: gb\_estli.\*  
10: gb\_est2.\*  
11: gb\_hcc.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24.2	78.1	459	10	BI221953
2	24.2	78.1	555	10	BI103096
3	22.6	72.9	197	9	AW229748
4	22.6	72.9	586	10	BI737971
5	21.2	68.4	463	12	AQ664401
c 6	21.2	67.7	658	12	AG146871
c 7	21.2	67.7	966	10	BQ292991
c 8	20.6	66.5	452	12	AQ213776
c 9	20.6	66.5	930	10	BF683668
10	20.4	65.8	331	10	BE708956
11	20.4	65.8	340	10	BE696387
c 12	20.4	65.8	357	10	BF925749
c 13	20.4	65.8	369	10	BF195626
14	20.4	65.8	383	10	BE745194
15	20.4	65.8	397	10	BG876258
c 16	20.4	65.8	401	10	BF881067
c 17	20.4	65.8	432	10	W78212

18	20.4	65.8	445	10	W30739
19	20.4	65.8	447	9	AW408381
20	20.4	65.8	458	9	AA812116
21	20.4	65.8	471	10	W93251
c 22	20.4	65.8	492	9	AW406526
23	20.4	65.8	496	10	BG385465
24	20.4	65.8	521	9	AA662909
25	20.4	65.8	563	10	BG116019
26	20.4	65.8	568	10	BF219924
27	20.4	65.8	607	10	BE795177
c 28	20.4	65.8	650	9	AA071197
29	20.4	65.8	675	9	AW027138
30	20.4	65.8	711	10	BI335232
31	20.4	65.8	715	12	AG177881
32	20.4	65.8	726	10	BG392336
33	20.4	65.8	757	10	BE513262
34	20.4	65.8	761	10	BE881549
35	20.4	65.8	789	10	BI832624
36	20.4	65.8	848	9	AU142879
37	20.4	65.8	855	10	BM046549
38	20.4	65.8	891	10	BI760539
39	20.4	65.8	895	10	BE873926
40	20.4	65.8	898	10	BG165109
41	20.4	65.8	901	10	BE617436
42	20.4	65.8	901	10	BE784393
43	20.4	65.8	924	10	BG775094
44	20.4	65.8	941	10	BG695997
45	20.4	65.8	964	10	BE747898

## ALIGNMENTS

BI221953 459 bp mRNA linear EST 11-JUL-2001  
602938496F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:5101702 5',  
mRNA sequence.  
BI221953  
BI221953.1 GI:14675397  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 459)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM11244 row: p column: 23  
High quality sequence start: 7  
High quality sequence stop: 459.  
Location/Qualifiers  
1. 459  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5101702"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

## FEATURES

source



found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM11927 row: 9 column: 14

High quality sequence stop: 586.

## FEATURES

## source

1. 586

/db\_xref="taxon:10090"

/clone="IMAGE:5364517"

/clone\_lib="NIH\_MGC\_94"

/tissue\_type="retina"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pCMV-SPORT6; Site:1: NotI;

Site:2: SalI; Cloned unidirectionally; oligo-dr primed.

Average insert size 3.3 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC Library."

115 a 167 c 180 g 124 t

BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 72.9%; Score 22.6; DB 10; Length 586;

Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cctctagagccaggatgcattgtgccc 30

Db 221 CCTCTACCTCCAGGATGCATTGTGGCC 249

## RESULT 5

## A0664401

## LOCUS

DEFINITION HS\_5480\_B1\_B09\_SP6E RPCI-11 Human Male BAC Library Homo sapiens

genomic clone Plate=1056 Col=17 Row=D, DNA sequence.

## ACCESSION

VERSION A0664401.1 GI:5172169

## KEYWORDS

SOURCE GSS;

## ORGANISM

human.

## REFERENCE

AUTHORS

1 (bases 1 to 463)

Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,

Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and

Hood L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server:

http://www.htsc.washington.edu

Plate: 1056 row: D column: 17

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 463.

Location/Qualifiers

1. 463

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=1056 Col=17 Row=D"

/clone\_lib="RPCI-11 Human Male BAC Library"

## FEATURES

## source

1. 463

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=1056 Col=17 Row=D"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site:1: EcoRI; Site:2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI Methylase. Size selected DNA was cloned into the

pBACE3.6 vector at EcoRI sites"

165 a 98 c 86 g 113 t 1 others

BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 68.4%; Score 21.2; DB 12; Length 463;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cctctagagccaggatgcattgttgc 29

Db 269 CACTTGACCCAGCATGATGTGTC 295

## RESULT 6

## AG146871/c

## LOCUS

DEFINITION Pan troglodytes DNA, clone: RP43-008J16.T7, genomic survey

sequence.

ACCESSION AG146871.1 GI:16676549

VERSION GSS; GSS (genome survey sequence).

KEYWORDS Pan troglodytes male lymphocytes DNA, clone\_lib:RPCI-43 Chimpanzee

SOURCE Male BAC Library clone:RP43-008J16.T7.

ORGANISM Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

1 (sites)

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

BAC end sequences of Library RPCI-43

Unpublished

2 (bases 1 to 658)

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:chimpesgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library RPCI-43 This BAC

end was generated during the R&D process and may have higher chance

of clone tracking errors.

PRIMERS

Sequencing: T7

LIBRARY

Vector : pBACE3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI.

Location/Qualifiers

1. 658

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="RP43-008J16.T7"

/sex="male"

/cell\_type="lymphocytes"

/clone\_lib="RPCI-43 Chimpanzee Male BAC Library"

230 a 132 c 107 g 189 t

BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 67.7%; Score 21; DB 12; Length 658;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cctctagagccaggatgcattgttgc 30

||||| |||||||| || |||||||| ||

```

Db      608  CCTAAGACCCAGCAATTAATTGTTTCC 580

RESULT  7
BG292991/c
LOCUS   966 bp      mRNA      linear      EST 21-FEB-2001
DEFINITION  602389672F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4501864 5',
mRNA sequence.
ACCESSION  BG292991
VERSION    BG292991.1  GI:13052346
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 966)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: The Cepko Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10369 row: o column: 17
           High quality sequence stop: 645.
           Location/Qualifiers
             1..966
               /organism="Mus musculus"
               /db_xref="taxon:10090"
               /clone="IMAGE:4501864"
               /clone_lib="NIH_MGC_94"
               /tissue_type="retina"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI;
               Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
               Average insert size 3.3 kb. Library enriched for
               full-length clones and constructed by Life Technologies.
               -Note: this is a NIH_MGC Library."

BASE COUNT  296 a  237 c  201 g  232 t
ORIGIN

Query Match      67.7%; Score 21; DB 10; Length 966;
Best Local Similarity 82.8%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY  3  ctctagagccaggatgcattgtgtgcc 31
||||| ||||||| ||||||| ||||||| |||
Db  659  CTCTAAGACCCAGGATGCATGTTGGCGC 631

RESULT  8
AQ213776
LOCUS   452 bp      DNA      linear      GSS 18-SEP-1998
DEFINITION  HS.2216 A2.D05.MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2216 Col=10 Row=G, DNA sequence.
ACCESSION  AQ213776
VERSION    AQ213776.1  GI:3624977
KEYWORDS  GSS.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 452)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
           Kellier,A., Shaker,K., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
           Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
           scanning the human genome
           Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
           99380589
           Contact: Mahairas GG, Wallace JC, Hood L
           High Throughput Sequencing Center
           University of Washington
           401 Queen Anne Avenue North, Seattle, WA 98109, USA
           Tel: (206) 616-3618
           Fax: (206) 616-3887
           Email: jwallace@u.washington.edu
           Sequence Tagged Connector
           Plate: 2216 row: G column: 10
           Class: BAC ends
           High quality sequence stop: 452.
           Location/Qualifiers
             1..452
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="Plate=2216 Col=10 Row=G"
               /clone_lib="CIT Approved Human Genomic Sperm Library D"
               /sex="male"
               /note="Organ: sperm; Vector: pBelOBAcl1; BAC Clones in
               E-Coli DH10B"

BASE COUNT  146 a  93 c  97 g  114 t  2 others
ORIGIN

Query Match      66.5%; Score 20.6; DB 12; Length 452;
Best Local Similarity 85.2%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  3  ctctagagccaggatgcattgtgtgc 29
||||| ||||||| ||||||| ||||||| |||
Db  401  CTCTTAGCCCCAGCATAGTTGC 427

RESULT  9
BF683668/c
LOCUS   930 bp      mRNA      linear      EST 23-DEC-2000
DEFINITION  60213991F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4301162 5',
mRNA sequence.
ACCESSION  BF683668
VERSION    BF683668.1  GI:11969076
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 930)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LICM1161 row: e column: 03
           High quality sequence stop: 445.
           Location/Qualifiers
             1..930
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:4301162"
               /clone_lib="NIH_MGC_46"
               /tissue_type="leiomyosarcoma cell line"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
               EcoRI; cDNA made by oligo-dT priming. Directionally cloned
```

into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 244 a 331 c 250 g 105 t

Query Match 66.5%; Score 20.6; DB 10; Length 930;  
Best Local Similarity 85.2%; Pred. No. 2.8e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cctctagagccagcagtcattgttgg 28

Db 376 CCTCTAGTCCAGGCTCCATTTGTTG 350

# RESULT 10

BE708956

LOCUS

DEFINITION QV2-H70577-160500-218-h08 HT0577 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE708956

VERSION BE708956.1 GI:10097221

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS

1 (bases 1 to 331)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV2-HT0577-160  
500-218-h08&st3=2000-05-16&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 56

High quality sequence stop: 197.

FEATURES

source

1. .331  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0577"  
/dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT

ORIGIN

75 a 79 c 91 g 86 t

Query Match 65.8%; Score 20.4; DB 10; Length 331;  
Best Local Similarity 80.0%; Pred. No. 2.6e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gctctagagccagcagtcattgttggc 30

Db 176 GGCTCAGAACCCAGGAGAAATTTGTC 205

# RESULT 11

BE696387

LOCUS

DEFINITION QV0-CT0383-230600-268-c04 CT0383 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE696387

VERSION BE696387.1 GI:10083547

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS

1 (bases 1 to 340)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV0-CT0383-230  
600-268-c04&st3=2000-06-23&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 9

High quality sequence stop: 338.

FEATURES

source

1. .340  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0383"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site:1: SmaI; Site:2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
,716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT

ORIGIN

93 a 81 c 89 g 77 t

Query Match

Best Local Similarity

Matches

24; Conservative

0; Mismatches

6; Indels

0; Gaps

65.8%; Score 20.4; DB 10; Length 340;

80.0%; Pred. No. 2.6e+02;

0; Mismatches

6; Indels

0; Gaps

0;

QY 1 gctctagagccagcagtcattgttggc 30

Db 172 GGCTCAGAACCCAGGAGAAATTTGTC 201





## Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM44t2=CM4-BT0859-011100-402-b09&t3=2000-11-01&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 31

High quality sequence stop: 100.

## FEATURES

source

1..383  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BT0859"  
/dev\_stage="Adult"

/note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 96 a 104 c 99 g 84 t

ORIGIN

Query Match 65.8%; Score 20.4; DB 10; Length 383;

Best Local Similarity 80.0%; Pred. No. 2.7e+02;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gctcttagagcccgagtcattgttgc 30

||||| ||||| ||||| ||||| |||||

Db 72 GCTCCTAGACCCAGGAGAAATTTGTTGCC 101

## RESULT 15

BG876258

LOCUS

QV0-CT0383-150200-116-c06 CT0383 Homo sapiens CDNA, mRNA sequence. EST 30-MAY-2001

ACCESSION

BG876258

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 397)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL

MEDLINE

COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=QV0-CT0383-150200-116-c06&t3=2000-02-15&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 5

High quality sequence stop: 397.

## FEATURES

source

Location/Qualifiers

1..397

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="CT0383"

/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 109 a 100 c 103 g 85 t

ORIGIN

Query Match 65.8%; Score 20.4; DB 10; Length 397;

Best Local Similarity 80.0%; Pred. No. 2.7e+02;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gctcttagagcccgagtcattgttgc 30

||||| ||||| ||||| ||||| |||||

Db 167 GGCTCCAGACCCAGGAGAAATTTGTTGCC 196

Search completed: July 15, 2002, 21:49:15

Job time: 18906 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:29:29 ; Search time 2368.24 Seconds  
(without alignments)  
335.780 Million cell updates/sec

Title: US-10-053-641-7  
Perfect score: 38  
Sequence: 1 ggctctagagcgagaaatcaatcggtactgagga 38

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pi.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_hgt\_hum.\*
- 31: em\_hgt\_inv.\*
- 32: em\_hgt\_other.\*
- 33: em\_hgtgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
-----					

1	30	78.9	59	6	I63542
2	30	78.9	59	6	I63546
3	30	78.9	128	6	I41185
4	30	78.9	226	12	SYNHIR
5	30	78.9	235	6	I41187
6	30	78.9	235	6	I41188
7	30	78.9	726	6	I41189
8	30	78.9	726	6	I41190
9	28.4	74.7	59	6	I63548
10	28	73.7	195	6	I25751
11	28	73.7	215	6	A36136
12	28	73.7	215	6	E08760
13	26.8	70.5	563	6	I25747
14	26.8	70.5	567	6	A36139
15	25.8	67.9	106	6	I41184
16	25.2	66.3	210	6	E02086
17	25.2	66.3	279	6	A01139
18	25.2	66.3	279	6	I26534
19	25.2	66.3	526	6	E02241
20	25.2	66.3	526	6	E03810
21	24.4	64.2	201	6	A34242
22	24.2	63.7	43	6	A30667
23	24.2	63.7	43	6	I38856
24	23.6	62.1	35	6	A30668
25	23.6	62.1	35	6	I38857
26	23.6	62.1	59	6	I63544
27	23.6	62.1	67	6	A34230
28	23.6	62.1	122	6	A34236
29	23.6	62.1	122	6	A34237
30	23.6	62.1	195	6	A03693
31	23.6	62.1	195	6	A03694
32	23.6	62.1	195	6	A43186
33	23.6	62.1	195	6	I62835
34	23.6	62.1	212	6	A13382
35	23.6	62.1	212	6	A13383
36	23.6	62.1	212	6	A34619
37	23.6	62.1	212	6	A34620
38	23.6	62.1	212	6	E00711
39	23.6	62.1	217	6	A03695
40	23.6	62.1	217	6	A03696
41	23.6	62.1	217	6	A34238
42	23.6	62.1	217	6	A34239
43	23.6	62.1	217	6	E00657
44	23.6	62.1	8491	6	AR031529
45	23.4	61.6	73174	2	AC036161

## ALIGNMENTS

RESULT	1				
LOCUS	I63542	I63542	59 bp	DNA	linear
DEFINITION	Sequence 5 from patent US 5663141.				
ACCESSION	I63542				
VERSION	I63542.1	GI:2481115			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 59)				
AUTHORS	Kurfuerst, M., Ruebsamen, K., Schmied, B., Koerwer, W., Schweden, J. and Hoeftken, H. Wolfgang.				
TITLE	Hirudin/polyalkylene glycol conjugates and hirudin muteins				
JOURNAL	Patent: US 5663141-A 5 02-SEP-1997;				
FEATURES	Location/Qualifiers				
source	1..59				
BASE COUNT	19 a	14 c	16 g	10 t	
ORIGIN					
Query Match		78.9%	Score 30;	DB 6;	Length 59;

Best Local Similarity 86.8%; Pred. No. 0.052; Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||||| || ||||||| |||||

Db 16 GGCTCTGACGGCGAAAAAACCCAGTGGTACTGGCGA 53

RESULT 2  
LOCUS I63546 59 bp DNA PAT 07:OCT-1997  
DEFINITION Sequence 9 from patent US 5663141.  
ACCESSION I63546  
VERSION I63546.1 GI:2481119  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 59)  
AUTHORS Kurfuerst, M., Ruebsaamen, K., Schmied, B., Koerwer, W., Schweden, J. and Hoeffken, H. Wolfgang.  
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins  
JOURNAL Patent: US 5663141-A 9 02-SEP-1997;  
FEATURES  
source Location/Qualifiers  
1..59  
/organism="unknown"

BASE COUNT 20 a 14 c 15 g 10 t  
ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 59;  
Best Local Similarity 86.8%; Pred. No. 0.052; Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||| ||||| ||||| ||||| |||||

Db 16 GGCTCTAAAGGCCAGAAAAACCCAGTGGTACTGGCGA 53

RESULT 3  
LOCUS I41185 128 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 5 from patent US 5624822.  
ACCESSION I41185  
VERSION I41185.1 GI:2081775  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 128)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 5 29-APR-1997;  
FEATURES  
source Location/Qualifiers  
1..128  
/organism="unknown"

BASE COUNT 20 a 33 c 34 g 41 t  
ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 128;  
Best Local Similarity 86.8%; Pred. No. 0.054; Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||||| ||||| ||||| ||||| |||||

Db 116 GGCTCTGACGGCGAAAAAACCCAGTGGTACTGGCGA 79

RESULT 4  
LOCUS SYNHIR 226 bp DNA SYN 27-APR-1993  
DEFINITION Synthetic hirudin gene, complete cds.

ACCESSION M26762  
VERSION M26762.1 GI:208478  
KEYWORDS hirudin; proteinase inhibitor; thrombin inhibitor.  
SOURCE Synthetic DNA.  
ORGANISM Synthetic construct  
artificial sequence.  
REFERENCE 1 (bases 1 to 226)  
AUTHORS Bergmann, C., Drott, J., Koehler, S., Fink, E. and Gassen, H. G.  
TITLE Chemical synthesis and expression of a gene coding for hirudin, the thrombin-specific inhibitor from the leech Hirudo medicinalis  
JOURNAL Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)  
MEDLINE 87026239  
FEATURES  
source Location/Qualifiers  
1..226  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
22..222  
/note="hirudin"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AAA72772.1"  
/db\_xref="GI:208478"  
/translation="MVFYTDCTESGNLCLCEGSNVGOGNKCILGSDGKNOQCVTGE  
GTFKPSHNDGDFEIEPEYLQ"

BASE COUNT 63 a 56 c 61 g 46 t  
ORIGIN

Query Match 78.9%; Score 30; DB 12; Length 226;  
Best Local Similarity 86.8%; Pred. No. 0.056; Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||||| ||||| ||||| ||||| |||||

Db 115 GGCTCTGACGGCGAAAAAACCCAGTGGTACTGGCGA 152

RESULT 5  
LOCUS I41187 235 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 7 from patent US 5624822.  
ACCESSION I41187  
VERSION I41187.1 GI:2081777  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 235)  
AUTHORS Koerwer, W.  
TITLE Hirudin fusion proteins and preparation of hirudin  
JOURNAL Patent: US 5624822-A 7 29-APR-1997;  
FEATURES  
source Location/Qualifiers  
1..235  
/organism="unknown"

BASE COUNT 72 a 63 c 57 g 43 t  
ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 235;  
Best Local Similarity 86.8%; Pred. No. 0.056; Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38  
||||| ||||||| ||||| ||||| ||||| |||||

Db 124 GGCTCTGACGGCGAAAAAACCCAGTGGTACTGGCGA 161

RESULT 6  
LOCUS I41188/c 235 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 8 from patent US 5624822.  
ACCESSION I41188  
VERSION I41188.1 GI:2081778

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 235)

AUTHORS Koerwer, W.

TITLE Hirudin fusion proteins and preparation of hirudin

JOURNAL Patent: US 5624822-A 8 29-APR-1997;

FEATURES Location/Qualifiers

1..235

source /organism="unknown"

BASE COUNT 42 a 58 c 64 g 71 t

ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 235;  
Best Local Similarity 86.8%; Pred. No. 0.056;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38

Db 116 GGCTCTGACGGCGAAAAAACCAACGTCGTTACTGGCGA 79

RESULT 7

LOCUS I41189

DEFINITION Sequence 9 from patent US 5624822.

ACCESSION I41189

VERSION I41189.1 GI:2081779

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 726)

AUTHORS Koerwer, W.

TITLE Hirudin fusion proteins and preparation of hirudin

JOURNAL Patent: US 5624822-A 9 29-APR-1997;

FEATURES Location/Qualifiers

1..726

source /organism="unknown"

BASE COUNT 281 a 171 c 126 g 148 t

ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 726;  
Best Local Similarity 86.8%; Pred. No. 0.059;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38

Db 616 GGCTCTGACGGCGAAAAAACCAACGTCGTTACTGGCGA 653

RESULT 8

LOCUS I41190/c

DEFINITION Sequence 10 from patent US 5624822.

ACCESSION I41190

VERSION I41190.1 GI:2081780

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 726)

AUTHORS Koerwer, W.

TITLE Hirudin fusion proteins and preparation of hirudin

JOURNAL Patent: US 5624822-A 10 29-APR-1997;

FEATURES Location/Qualifiers

1..726

source /organism="unknown"

BASE COUNT 149 a 126 c 171 g 280 t

ORIGIN

Query Match 78.9%; Score 30; DB 6; Length 726;  
Best Local Similarity 86.8%; Pred. No. 0.059;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38

Db 111 GGCTCTGACGGCGAAAAAACCAACGTCGTTACTGGCGA 74

RESULT 9

LOCUS I63548

DEFINITION Sequence 11 from patent US 5663141.

ACCESSION I63548

VERSION I63548.1 GI:2481121

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 59)

AUTHORS Kurfuerst, M., Ruebsamen, K., Schmied, B., Koerwer, W., Schweden, J. and

Hoeffken, H. Wolfgang.

TITLE Hirudin/polyalkylene glycol conjugates and hirudin mutants

JOURNAL Patent: US 5663141-A 11 02-SEP-1997;

FEATURES Location/Qualifiers

1..59

source /organism="unknown"

BASE COUNT 18 a 14 c 16 g 11 t

ORIGIN

Query Match 74.7%; Score 28.4; DB 6; Length 59;  
Best Local Similarity 84.2%; Pred. No. 0.24;  
Matches 32; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38

Db 16 GGCTCTAAGCGGAACGTAACCAACGTCGTTACTGGCGA 53

RESULT 10

LOCUS I25751

DEFINITION Sequence 27 from patent US 5552299.

ACCESSION I25751

VERSION I25751.1 GI:1605621

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 195)

AUTHORS Ott, Ian., Klupp, T., Moln ar, Ian., Pathy, Aas., Barta, Ian., Bark o n

ee T oth, Z., Ambrus, Gabor., Sal at, Janos., Tegdes, Ao.,

Moravcsik, I., Egy ud, C., Albrecht, Karnly., K oncz O., Kalmán.,

Vincze, A., Barab as, E., M at e, Gorgy., Kiss, Gorgy. B., Kiss, Peter.,

P olya, Kálmán., Erdei, Janos., Guly as, E. and Zilahi, E.

Plasmids and process for producing recombinant desulphatohirudin

HV-1 peptides

JOURNAL Patent: US 5552299-A 27 03-SEP-1996;

FEATURES Location/Qualifiers

1..195

source /organism="unknown"

BASE COUNT 58 a 40 c 45 g 51 t 1 others

ORIGIN

Query Match 73.7%; Score 28; DB 6; Length 195;  
Best Local Similarity 81.6%; Pred. No. 0.37;  
Matches 31; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcggttactggcga 38

RESULT 14

Search completed: July 15, 2002, 22:29:30  
Job time: 19016 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:55 ; Search time 758.37 Seconds  
(without alignments)  
86.030 Million cell updates/sec

Title: US-10-053-641-7  
Perfect score: 38  
Sequence: 1 ggctctagaggcgaaataatcaatgctgactgscga 38

Scoring table: IDENTITY\_NUC  
Gap 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	78.9	726	12	AAQ12379
2	28.4	74.7	468	16	AAQ93244
3	28	73.7	219	15	AAQ54995
4	26.8	70.5	567	15	AAQ54997
5	25.2	66.3	195	13	AAQ27601
6	25.2	66.3	208	10	AAQ91851
7	25.2	66.3	276	13	AAQ27600
8	25.2	66.3	279	12	AAQ13831
9	25.2	66.3	525	11	AAQ06873
					Hirudin peptide/Protein A fusion gene.
					anticoagulant; fusion protein; ds.
					Location/Qualifiers
					mat_peptide
					1.489
					/tag= a
					/product= Protein A
					529..720
					/tag= b
					/product= Hirudin
					490..528
					/tag= c
					/product= linker oligopeptide Y
					DB3942580-A.
					27-JUN-1991.
					22-DEC-1989;
					89DE-3942580.
					22-DEC-1989;
					89DE-3942580.
					(BADI ) BASF AG.

10	25.2	66.3	526	11	AAQ03237	Sequence of DNA in
11	24.4	64.2	231	7	AAQ60354	Desulphatohirudin
12	24.2	63.7	43	14	AAQ43368	Hirudin oligonucle
13	23.6	62.1	35	14	AAQ43369	Hirudin oligonucle
14	23.6	62.1	195	16	AAQ81528	Leech hirudin mute
15	23.6	62.1	196	8	AAQ70319	Sequence encoding
16	23.6	62.1	217	7	AAQ60355	Desulphatohirudin
17	23.6	62.1	217	8	AAQ70323	Sequence of the de
18	23.6	62.1	227	22	AAQ61507	S. marcescens hiru
19	22.2	58.4	2445	19	AAQ30596	H. pylori outer me
20	22.2	58.4	2445	20	AAQ75781	H. pylori outer me
21	22.2	58.4	2448	19	AAQ30595	H. pylori outer me
22	22.2	58.4	2448	20	AAQ75780	Partial Hirudin HV
23	22	57.9	183	12	AAQ14927	Synthetic hirudin
24	22	57.9	185	12	AAQ14926	Sequence encoding
25	22	57.9	198	13	AAQ24986	Hirudin variant.
26	22	57.9	198	15	AAQ63876	HV-1 gene. Synthe
27	22	57.9	210	12	AAQ10452	DNA encoding hirud
28	22	57.9	224	7	AAQ60746	Sequence encoding
29	22	57.9	229	9	AAQ81294	Sequence encoding
30	22	57.9	238	9	AAQ81295	Synthetic HVI gene
31	22	57.9	257	10	AAQ91836	Synthetic hirudin
32	22	57.9	257	10	AAQ91867	CUP1 promoter, PHO
33	22	57.9	358	6	AAQ50397	Yeast CUP1 promote
34	22	57.9	1082	15	AAQ64146	PJDB207/GAPFL-YHIR
35	22	57.9	1082	15	AAQ45312	Salmonella typhi D
36	22	57.9	1130	16	AAQ81527	Human CDNA encodin
37	21	55.3	1773	23	AAQ56019	DNA encoding novel
38	20.8	54.7	745	22	AAQ33871	Synthetic hirudin
39	20.6	54.2	3884	23	AAQ90132	Hirudin HV-1. Syn
40	20.4	53.7	201	12	AAQ12153	Desulphatohirudin
41	20.4	53.7	201	13	AAQ25184	Factor Xa-cleavabl
42	20.4	53.7	304	15	AAQ54996	Factor Xa-cleavabl
43	20.4	53.7	420	12	AAQ12155	Factor Xa-cleavabl
44	20.4	53.7	1458	12	AAQ12162	Factor Xa-cleavabl
45	20.4	53.7	1467	12	AAQ12490	

ALIGNMENTS

RESULT 1	
AAQ12379	
ID	AAQ12379 standard; DNA; 726 BP.
XX	
AC	AAQ12379;
XX	
DT	17-SEP-1991 (first entry)
XX	
DE	Hirudin peptide/Protein A fusion gene.
XX	
KW	anticoagulant; fusion protein; ds.
XX	
PH	Key
FT	mat_peptide
FT	1.489
FT	/tag= a
FT	/product= Protein A
FT	529..720
FT	/tag= b
FT	/product= Hirudin
FT	490..528
FT	/tag= c
FT	/product= linker oligopeptide Y
XX	
PN	DB3942580-A.
XX	
PD	27-JUN-1991.
XX	
PF	22-DEC-1989;
XX	
PR	89DE-3942580.
XX	
PA	22-DEC-1989;
	89DE-3942580.
	(BADI ) BASF AG.

```

XX PI Korwer W;
XX XX
XX DR WPI; 1991-194236/27.
XX DR P-PSDB; AAR12751.
XX XX
XX PT Hirudin peptide prodn. by cleaving new fusion peptide - of
XX PT hirudin and protein A, expressed in high yield and stable,
XX PT soluble form by transformed E.coli
XX XX
XX PS Example 1; Page 6-7; 9pp; German.
XX CC This sequence is an example of a fusion construct for expression of
XX CC the fusion peptide of the invention. The Met residue in the linker
XX CC oligopeptide allows cleavage by CNBr to release two fragments which
XX CC can be easily separated by Igc affinity chromatography. (The protein
XX CC A component binds to Igc sepharose). Increased yields of hirudin are
XX CC obtained using this fusion construct.
XX XX
XX SQ Sequence 726 BP; 280 A; 171 C; 126 G; 149 T; 0 other;

Query Match 78.9%; Score 30; DB 12; Length 726;
Best Local Similarity 86.8%; Pred. No. 0.0037;
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctctagaggcgagaaatacaatcgcttactggcgga 38
||||| ||||||| || |||||||
DB 616 ggctctagaggcgagaaatacaatcgcttactggcgga 653

RESULT 2
AAQ93244
ID AAQ93244 standard; DNA; 468 BP.
XX AC AAQ93244;
XX XX
XX DT 01-NOV-1995 (first entry)
XX DE
XX DE Fusion construct of glucoamylase-hirudin DNA.
XX KW expression cassette; recombinant protein; production; Hansenula;
XX KW yeast; leader sequence; adaptor; alpha helix; glucoamylase;
XX KW secretion; processing; thrombin inhibitor; hirudin; ds.
XX OS Synthetic.
XX XX
XX FH Key Location/Qualifiers
XX FT primer_bind complement (1..33)
XX FT /tag= a
XX FT /note= "primer AAQ85840 binding site"
XX FT 13..444
XX FT /tag= b
XX FT /product= glucoamylase-hirudin fusion protein
XX FT 199..223
XX FT /tag= c
XX FT /note= "primer AAQ85841 binding site"
XX FT complement (220..267)
XX FT /tag= d
XX FT /note= "primer AAQ85842 binding site"
XX FT 13..228
XX FT /tag= e
XX FT /note= "encodes amino acids 1-72 of glucoamylase"
XX FT 241..246
XX FT /tag= f
XX FT /note= "encodes processor signal"
XX FT 247..444
XX FT /tag= g
XX FT /note= "encodes hirudin-h120"
XX FT 0..468
XX FT /tag= h
XX FT /note= "primer AAQ85843 binding site"
XX FT
XX XX

```

```

PN DE4329969-A.
XX XX
XX PD 09-MAR-1995.
XX XX
XX PF 04-SEP-1993; 93DE-4329969.
XX XX
XX PR 04-SEP-1993; 93DE-4329969.
XX XX
XX PA (BADI ) BASF AG.
XX PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX PI Bollsweiler C, Janowicz ZA, Plontek M, Schweden J;
XX PI Strasserawm, Weydemann U;
XX XX
XX DR WPI; 1995-107810/15.
XX DR P-PSDB; AAR76951.
XX XX
XX PT Recombinant protein prodn. in Hansenula yeast - transformed with
XX PT expression cassette contg. leader, adaptor, processing signal and
XX PT gene, provides efficient secretion and correct processing
XX PS Example 1; Fig 1; 10pp; German.
XX XX
XX CC AAQ93244 is a DNA fusion product resulting from ligation of an
XX CC EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding
XX CC a leader sequence, an adaptor (see AAR71472, amino acids 23-72 of GAM
XX CC (glucamylase from Schwanniomycetes occidentalis, plus -His-Pro-Leu-Gln at
XX CC the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,
XX CC encoding hirudin. AAQ93244 is an example of an expression vector insert
XX CC of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;
XX CC P = processor; Gene = structural gene). The cassettes ensure efficient
XX CC secretion and correct processing of heterologous structural genes in
XX CC yeast of the genus Hansenula, and so provides high yields of mature
XX CC proteins and facilitates subsequent purification.
XX SQ Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match 74.7%; Score 28.4; DB 16; Length 468;
Best Local Similarity 84.2%; Pred. No. 0.016; Mismatches 6; Indels 0; Gaps 0;
Matches 32; Conservative 0;

QY 1 ggctctagaggcgagaaatacaatcgcttactggcgga 38
||||| ||||||| || |||||||
DB 337 ggctctagaggcgagaaatacaatcgcttactggcgga 374

RESULT 3
AAQ54995
ID AAQ54995 standard; DNA; 219 BP.
XX AC AAQ54995;
XX XX
XX DT 11-JUL-1994 (first entry)
XX DE
XX DE Desulphatohirudin gene with E. coli-Saccharomyces codon usage.
XX KW Hirudin; HV-1; E. coli; codon usage; desulphatohirudin; HV01; 33Asp;
XX KW 33 Asp; biological activity; thrombosis; thromboembolism; ds.
XX OS Hirudo medicinalis.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 14..208
XX FT /tag= a
XX FT /product= Hirudin_HV-1
XX XX
XX EP576792-A.
XX XX
XX XX 05-JAN-1994.
XX XX
XX PF 13-APR-1993; 93EP-0105848.
XX XX

```

```

PR 09-APR-1992; 92HU-0001200.
PA (BIOG ) BIOGAL GYOGYSZERGYAR.
PI Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;
PI Botond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;
PI Mate G, Molnari, Moravcsik I, Ott I, Pathy A;
PI Poelya K, Salat J, Tegdes A, Vincze A, Zilahi E;
XX WPI: 1994-009153/02.
DR P-PSDB; AAR47488.
XX Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using
PT E. coli, Saccharomyces and Streptomyces hosts, for increased
PT yields
XX Disclosure; Page 44; 79pp; English.
XX This sequence represents the hirudin HV-1 gene designed on the
CC basis on E. coli codon usage. This sequence was used in the
CC production of desulphatohirudin HV01 33Asp and desulphatohirudin HV-1
CC 33 Asn. The expressed peptides produced using this sequence have the
CC same biological activity as natural hirudin and can be used in the
CC treatment of thrombosis, thromboembolism, etc. Using naturalised
CC sequences such as this, large amounts of hirudin may be produced
CC stably, with the highest production level achieved being 140-180
CC mg/litre of culture.
XX Sequence 219 BP; 63 A; 46 C; 51 G; 58 T; 1 other;
SQ
Query Match 73.7%; Score 28; DB 15; Length 219;
Best Local Similarity 81.6%; Pred. No. 0.021;
Matches 31; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 1 ggctctagagcgcaaaaaaatcaatgcgttactggcga 38
Db 104 ggtctcragcgtgaaaaaatcaatgtgtcactggcga 141
RESULT 4
AAQ54997
ID AAQ54997 standard; DNA; 567 BP.
XX AAQ54997;
AC AAQ54997;
XX 11-JUL-1994 (first entry)
XX Desulphatohirudin gene expression/excretion cassette.
XX Hirudin; HV-1; E. coli; codon usage; desulphatohirudin; HV01; 33Asp;
XX 33 Asn; biological activity; thrombosis; thromboembolism; ss.
XX Hirudo medicinalis.
XX Key Location/Qualifiers
FH -35_signal 43..48
FT /*tag= a
FT -10_signal 68..73
FT /*tag= b
FT misc_signal 83
FT /*tag= c
FT /note= "Transcription initiation site"
FT RBS 260..266
FT /*tag= d
FT /note= "Shine-Dalgarno sequence"
FT sig_peptide 269..352
FT /tag= e
FT mat_peptide 353..550
FT /*tag= f
FT /product= Hirudin
XX EP576792-A.

```

```

XX 05-JAN-1994.
XX 13-APR-1993; 93EP-0105848.
XX 09-APR-1992; 92HU-0001200.
XX (BIOG ) BIOGAL GYOGYSZERGYAR.
XX Albrecht K, Ambrus G, Barabas E, Barta I, Berk T;
XX Botond K, Egyued C, Erdei J, Kiss P, Klupp T, Koenczoel K;
XX Mate G, Molnari, Moravcsik I, Ott I, Pathy A;
XX Poelya K, Salat J, Tegdes A, Vincze A, Zilahi E;
XX WPI: 1994-009153/02.
DR P-PSDB; AAR47490.
XX Prodn. of recombinant desulphatohirudin HV-1 peptide(s) - using
PT E. coli, Saccharomyces and Streptomyces hosts, for increased
PT yields
XX Disclosure; Page 48; 79pp; English.
XX This sequence represents the an expression/secretion cassette for
CC the expression of hirudin HV-1. This sequence was used in the
CC production of desulphatohirudin HV01 33Asp and desulphatohirudin HV-1
CC 33 Asn. The expressed peptides produced using this sequence have the
CC same biological activity as natural hirudin and can be used in the
CC treatment of thrombosis, thromboembolism, etc. Using naturalised
CC sequences such as this, large amounts of hirudin may be produced
CC stably, with the highest production level achieved being 140-180
CC mg/litre of culture.
XX Sequence 567 BP; 162 A; 123 C; 134 G; 148 T; 0 other;
SQ
Query Match 70.5%; Score 26.8; DB 15; Length 567;
Best Local Similarity 81.6%; Pred. No. 0.08;
Matches 31; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 ggctctagagcgcaaaaaaatcaatgcgttactggcga 38
Db 443 ggtctcragcgtgaaaaaatcaatgtgtcactggcga 480
RESULT 5
AAQ27601
ID AAQ27601 standard; DNA; 195 BP.
XX AAQ27601;
AC AAQ27601;
XX 04-FEB-1993 (first entry)
XX Encodes hirudin derived thrombin inhibitor #2.
XX bleeding time; clotting time; anti-thrombotic;
XX congestive phlebothrombosis.
XX Hirudo medicinalis.
XX JP04197184-A.
XX 16-JUL-1992.
XX 28-NOV-1990; 90JP-0323133.
XX 28-NOV-1990; 90JP-0323133.
XX (MITK ) MITSUI TOATSU CHEM INC.
XX WPI: 1992-288441/35.
XX P-PSDB; AAR26314.

```

PT Thrombin-inhibiting polypeptide(s) more potent than hirudin HV1 -  
 PT produced by transformed Bacillus Subtilis, for treating  
 PT thrombosis

XX Claim 6; Page 1; 24pp; Japanese.

XX This sequence encodes a novel, hirudin derived thrombin inhibitor.

XX Sequence 195 BP; 76 A; 24 C; 42 G; 53 T; 0 other;

Query Match 66.3%; Score 25.2; DB 13; Length 195;  
 Best Local Similarity 78.9%; Pred. No. 0.31;  
 Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgagaaataatcaatgcgttactggcga 38

Db 91 ggatctgatggagagaaataatcaatggttacaggaga 128

# RESULT 6

AAQ1851  
 ID AAQ1851 standard; DNA; 208 BP.

XX AC AAQ1851;

XX 26-MAR-1990 (first entry)

XX Synthetic DNA encoding desulphatohirudin.

XX Desulphatohirudin; synthetic gene; thrombin inhibitor; HV-1 gene;  
 KW plasmid p3010; plasmid p4014.

XX JP01247092-A.

XX 02-OCT-1989.

XX 29-MAR-1988; 88JP-0073200.

XX 29-MAR-1988; 88JP-0073200.

XX (MITK ) MITSUI TOATSU CHEM. INC.

XX WPI; 1989-330037/45.

XX Prepn. of desulphatohirudin protein used as thrombin inhibitor - by  
 PT introducing specific DNA into expression vector, inserting obtd.  
 PT recombinant plasmid into host and incubating.

XX Claim 1; page 485; 6pp; Japanese.

XX The synthetc gene is constructed by dividing the hirudin HV-1 gene into  
 CC 7 parts and ligating to form the gene. It is inserted into pBR322 to  
 CC give plasmid p3010. This is inserted into expression vector pKK223-3 to  
 CC give plasmid p4014. This is used to transform E.coli. It produces  
 CC desulphatohirudin, which is active as a thrombin inhibitor. It allows  
 CC easy mass prodn.

XX Sequence 208 BP; 82 A; 23 C; 43 G; 60 T; 0 other;

Query Match 66.3%; Score 25.2; DB 10; Length 208;  
 Best Local Similarity 78.9%; Pred. No. 0.31;  
 Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgagaaataatcaatgcgttactggcga 38

Db 97 ggatctgatggagagaaataatcaatggttacaggaga 134

# RESULT 7

AAQ27600  
 ID AAQ27600 standard; DNA; 276 BP.

XX AAQ27600;  
 AC 04-FEB-1993 (first entry)  
 DT Encodes hirudin derived thombin inhibitor #1.  
 XX bleeding time; clotting time; anti-thrombotic;  
 KW congestive phlebothrombosis.  
 XX Hirudo medicinalis.

XX JP04197184-A.

XX 16-JUL-1992.

XX 28-NOV-1990; 90JP-0323133.

XX 28-NOV-1990; 90JP-0323133.

XX (MITK ) MITSUI TOATSU CHEM INC.

XX WPI; 1992-288441/35.

XX P-PSDB; AAR26313.

XX Thrombin-inhibiting polypeptide(s) more potent than hirudin HV1 -  
 PT produced by transformed Bacillus Subtilis, for treating  
 PT thrombosis

XX Claim 4; Page 1; 24pp; Japanese.

XX This sequence encodes a novel, hirudin derived thrombin inhibitor.

XX Sequence 276 BP; 91 A; 39 C; 66 G; 80 T; 0 other;

Query Match 66.3%; Score 25.2; DB 13; Length 276;  
 Best Local Similarity 78.9%; Pred. No. 0.33;  
 Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgagaaataatcaatgcgttactggcga 38

Db 172 ggatctgatggagagaaataatcaatggttacaggaga 209

# RESULT 8

AAQ13831  
 ID AAQ13831 standard; DNA; 279 BP.

XX AC AAQ13831;

XX 09-DEC-1991 (first entry)

XX MSP signal peptide-hirudin gene fusion.

XX Major secretion product; expression cassette; desulfatohirudin; ss.

XX Lactococcus lactis LM0230 (DSM 5805), Hirudo medicinalis.

XX Key Location/Qualifiers

FT sig\_peptide 1..81  
 FT /\*tag= a  
 FT /note= "MSP signal peptide"

FT mat\_peptide 82..279  
 FT /\*tag= b  
 FT /note= "hirudin structural gene."

XX EP449770-A.

XX 02-OCT-1991.

XX 13-MAR-1991; 91EP-0810167.

```

22-MAR-1990; 90GB-0006400.
(CIBA ) CIBA GEIGY AG.
Suri B, Schmitz A;
WPI: 1991-290162/40.
P-PSDB; AARI4151.
Hybrid vectors for expression of polypeptide(s) - comprise DNA
from plasmids obtd. from Lactococcus lactis, esp. major secretion
prod. gene fragments.
PS
PS Disclosure; Page 26; 36pp; English.
CC
CC The sequence (SEQ ID NO:2) was obtd. by ligation of the signal
CC peptide encoding DNA from L. lactis (from pUCRS, DSM 5803) and the
CC coding region for desulphatohirudin (from PML310, EP-168342). of
CC The presence of the signal peptide results in the secretion of
CC expressed hirudin into the supernatant.
CC See also AAQ13830.
XX
XX Sequence 279 BP; 76 A; 66 C; 68 G; 69 T; 0 other;
SQ
Query Match 66.3%; Score 25.2; DB 12; Length 279;
Best Local Similarity 78.9%; Pred. NO. 0.33;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 1 ggctctagagcgcgaaaaaatcaatgcgttactcggcga 38
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 ggctctgacgctgaaaaaacaccagtcggtaccgcga 209
RESULT 9
AAQ06873
ID AAQ06873 standard; DNA; 525 BP.
XX
XX AAQ06873;
DT 04-MAR-1991 (first entry)
DE Secretion plasmid for prodn. of HV1 hirudin.
XX
XX Thrombin inhibitor; ss.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH -35_signal 179..184
FT /*tag= a
FT -10_signal 203..208
FT /*tag= b
FT RBS 236..243
FT /*tag= c
FT sig_peptide 251..331
FT /*tag= d
FT /*label="secretion signal from Bacillus amyloliqui-
FT faciens"
FT misc_RNA 332..337
FT /*tag= e
FT /*product= dipeptide link
FT mat_peptide 338..525
FT /*tag= f
FT /*product= modified HV1-hirudin
XX
XX EP402159-A.
XX
XX 12-DEC-1990.
XX
XX 08-JUN-1990; 90EP-0306248.
XX

```



XX DB4140381-A.  
 XX 09-JUN-1993.  
 XX 07-DEC-1991; 91DE-4140381.  
 XX 07-DEC-1991; 91DE-4140381.  
 XX (FARH ) HOECHST AG.  
 XX Crause P, Habermann P, Schmid G, Tripiier D, Ulmer W;  
 XX WPI; 1993-189473/24.  
 XX New stable synthetic iso:hirudin(s) useful as medicaments -  
 PT having amino acid exchanges at positions 33,52,53,54 and 55, useful  
 PT in treatment of thrombosis.  
 XX Example 2; Page 4; 12pp; German.  
 XX Example 2 describes the prodn. of hirudin variants having Ala  
 CC as N-terminal. Hir5 and Hir6 are used in the prodn. of variant 13  
 CC (Ala1,Glul3,Glul52,Glul53,Glul55) (AAR37666).  
 CC The new isohirudin allows longer storage, and injectable solns. can  
 CC be produced and stored. There is no increase in antigenicity. The  
 CC N-terminus does not influence stability. Isohirudin may be useful  
 CC in thrombosis therapy.  
 XX Sequence 35 BP; 5 A; 12 C; 6 G; 12 T; 0 other;

Query Match 62.1%; Score 23.6; DB 14; Length 35;  
 Best Local Similarity 86.7%; Pred. No. 1.1;  
 Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 9 aggcgaataaaatcaatgcgtactgcga 38  
 ||||| || || ||||| |||||  
 Db 32 AGGTGAAGAAAGACCACTGCTTACTGCCGA 3

RESULT 14  
 AAQ81528  
 ID AAQ81528 standard; DNA; 195 BP.

XX AAQ81528;  
 XX 29-SEP-1995 (first entry)  
 XX Leech hirudin mutein 33-Cys HV1.

XX Hirudin mutant; leech; anticoagulant; antithrombotic; Cys 33 HV1; ss.  
 XX Synthetic.

Key Location/Qualifiers  
 CDS 1..195  
 FT /\*tag= a

XX WO9504823-A.

XX 16-FEB-1995.

XX 23-JUL-1994; 94WO-EP02438.

XX 04-AUG-1993; 93EP-0810552.

XX (CIBA ) CIBA GEIGY AG.

XX (UPCG-) UPC GEN-PHARMA AG.

XX Chang JY, Grossenbacher H, Marki W, Maerki W;

XX WPI; 1995-090901/12.

DR P-PSDB; AAR69100.

XX Conjugate consisting of 2-4 residues of desulphatohirudin muteins  
 PT - having hirudin activity, used in compns. to treat thromboses  
 PT and embolism

XX Claim 34; Page 38; 44pp; English.

XX The wild-type hirudin gene is contd. in a plasmid, pJDB207/GAPFL-  
 CC YHR, described in EP-A-340 170. It is used as a template in a PCR  
 CC to construct plasmid pDP34/GAPFL-YHR(CYS33), a mutant of the  
 CC recombinant desulphatohirudin derivative 1 (HV1) with Cys replacing  
 CC Asp at AA posn. 33. This results in the expression of  
 CC desulphatohirudin-Cys(33) (Cys(33)HV1). The primers used for the PCR  
 CC mutagenesis are AAQ81523-Q81526, primers 1-4 respectively. Primers 1  
 CC and 4 are the upstream and downstream primers for amplification of  
 CC the template DNA; primers 2 and 3 are the internal primers which are  
 CC partly complementary to each other and both cover the site of  
 CC mutation. The resulting plasmid is referred to as pDP34/GAPFL-YHR  
 CC (CYS33) (see AAQ81528/R69100). An expression cassette where the  
 CC desulphatohirudin mutein or deriv. is AAQ81528 is claimed, and  
 CC a desulphatohirudin HV1 wherein Asp33 is replaced by Cys (AAR69100)  
 XX is also claimed.

SQ Sequence 195 BP; 57 A; 38 C; 47 G; 53 T; 0 other;

Query Match 62.1%; Score 23.6; DB 16; Length 195;  
 Best Local Similarity 76.3%; Pred. No. 1.5;  
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 1 ggcctagagcgcaaaaaaatcaatgcgtactgcga 38  
 ||||| || || ||||| || ||||| |||||  
 Db 91 ggtctcttggtgaaagaacaaatggtgtaccggtga 128

RESULT 15  
 AAN70319

ID AAN70319 standard; DNA; 196 BP.

XX AAN70319;

XX 02-APR-1991 (first entry)

XX Sequence encoding desulphatohirudin variant 1 (HIV).

XX Anticoagulant; thrombin inhibitor; ds.

Key Location/Qualifiers  
 CDS 1..196  
 FT /\*tag= a

XX EP225633-A.

XX 16-JUN-1987.

XX 09-DEC-1986; 86EP-0117098.

XX 29-MAY-1986; 86GB-0013088.

XX 12-DEC-1985; 85GB-0030631.

XX (CIBA ) CIBA GEIGY AG.

XX (PLAN-) PLANTORGAN WERK HEINRICH.

XX (CHRI-) PLANTORGANW CHRISTENSEN.

XX Meyhack B, Marki W, Helm J;

XX WPI; 1987-164868/24.

XX P-PSDB; AAP70225.

XX New DNA constructs and hybrid vectors for transformation of yeast

PT etc. - useful for prodn. and secretion of protein with hirudin

PT activity for use as thrombin inhibitors.

XX Disclosure; p18; 146pp; English.  
XX  
CC The preferred DNA construct of the invention contains the PHO5  
CC promoter and a DNA segment consisting of the PHO5 signal sequence  
CC upstream of and in reading frame with a DNA sequence coding for  
CC mature desulphatohirudin. The segment is under the transcriptional  
CC control of the PHO5 promoter and the 3' flanking sequence of the  
CC PHO5 gene.  
XX  
SQ Sequence 196 BP; 54 A; 49 C; 52 G; 41 T; 0 other;  
  
Query Match 62.1%; Score 23.6; DB 8; Length 196;  
Best Local Similarity 76.3%; Pred. No. 1.5;  
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 1 ggctctagagcgcaaaaaaatcaatgcgttactggcga 38  
||| || ||||| || ||||| |||  
Db 91 ggtctgacggtgaaaaaacaccagtgcggtaccggtga 128

Search completed: July 15, 2002, 22:44:55  
Job time: 10068 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:31:51 ; Search time 174.99 seconds  
(without alignments)  
53.341 Million cell updates/sec

Title: US-10-053-641-7

Perfect score: 38

Sequence: 1 ggctctagagcgagaaataatcgcttactgcca 38

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/1/lna/6A\_COMB.seq:\*

4: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:\*

5: /cgn2\_6/ptodata/1/lna/PTUS\_COMB.seq:\*

6: /cgn2\_6/ptodata/1/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	78.9	59	1	US-07-859-453E-5
2	30	78.9	59	1	US-07-859-453E-9
c	3	30	78.9	128	1 US-08-262-384A-5
	4	30	78.9	235	1 US-08-262-384A-7
c	5	30	78.9	235	1 US-08-262-384A-8
	6	30	78.9	726	1 US-08-262-384A-9
c	7	30	78.9	726	1 US-08-262-384A-10
	8	28.4	74.7	59	1 US-07-859-453E-11
	9	28	73.7	195	1 US-08-044-506B-27
10	26.8	70.5	563	1	US-08-044-506B-22
11	25.8	67.9	106	1	US-08-262-384A-4
12	25.2	66.3	195	6	5166318-14
13	25.2	66.3	276	6	5166318-15
14	25.2	66.3	279	6	5166318-15
15	25.2	66.3	526	6	5166318-16
16	24.2	63.7	43	1	US-07-985-110-9
17	24.2	63.7	43	1	US-08-099-053-9
18	24.2	63.7	43	1	US-08-452-829-9
c	19	23.6	62.1	35	1 US-07-985-110-10
	20	23.6	62.1	35	1 US-08-099-053-10
c	21	23.6	62.1	35	1 US-08-452-829-10
	22	23.6	62.1	59	1 US-07-859-453E-7
23	23.6	62.1	67	6	5422249-8
24	23.6	62.1	122	6	5422249-9
25	23.6	62.1	195	1	US-08-406-948A-7
26	23.6	62.1	212	6	5180668-10
27	23.6	62.1	217	6	5422249-13

28	23.6	62.1	227	2	US-07-982-064-8	Sequence 8, Appli
c	29	23.6	8491	2	US-08-757-439-1	Sequence 1, Appli
	30	22	46	2	US-08-861-459-7	Sequence 7, Appli
	31	22	198	1	US-07-910-528-2	Sequence 2, Appli
	32	22	198	1	US-08-348-972-2	Sequence 2, Appli
	33	22	198	2	US-08-861-459-3	Sequence 3, Appli
	34	22	1082	1	US-08-715-252-1	Sequence 1, Appli
	35	22	57.9	1082	2 US-08-453-051-3	Sequence 3, Appli
	36	22	57.9	1130	1 US-08-406-948A-5	Sequence 5, Appli
c	37	20.6	38	2	US-08-861-459-9	Sequence 9, Appli
	38	20.6	54.2	45	1 US-08-044-506B-12	Sequence 12, Appli
	39	20.4	46	2	US-08-861-459-4	Sequence 4, Appli
	40	20.4	195	1	US-08-044-506B-28	Sequence 28, Appli
	41	20.4	201	1	US-07-854-596B-1	Sequence 1, Appli
	42	20.4	202	1	US-08-044-506B-20	Sequence 20, Appli
c	43	20.4	206	1	US-08-044-506B-21	Sequence 21, Appli
	44	20.4	223	1	US-07-854-596B-7	Sequence 7, Appli
	45	20.4	420	1	US-07-854-596B-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-07-859-453E-5  
; Sequence 5, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schweden, Juetgen  
; APPLICANT: Hoefken, Hans Wolfgang  
; TITLE OF INVENTION: Hlurudin/polyalkylene glycol  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS version 7.0  
; SOFTWARE: Wordperfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/859,453E  
; FILING DATE: 29-MAY-1992  
; CLASSIFICATION: 514  
; CLASSIFICATION: A 61 K 37/64  
; CLASSIFICATION: A 61 K 47/48  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/01998  
; FILING DATE: 22-NOV-1990  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 59 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; US-07-859-453E-5

Query Match 78.9%; Score 30; DB 1; Length 59;  
Best Local Similarity 86.8%; Pred. No. 0.0012;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 ggctctagagcgagaaataatcgcttactgcca 38  
||||| ||||||||| ||||||||| |||

Db 16 GGCTCTGACGGCGAAGAAAAAACCACTGCGTACTGGCGA 53

## RESULT 2

US-07-859-453E-9  
; Sequence 9, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schweden, Juergen  
; APPLICANT: Hoeftken, Hans Wolfgang  
; TITLE OF INVENTION: Hirudin/polyalkylene glycol  
; NUMBER OF INVENTION: conjugates  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS version 7.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/859,453E  
; FILING DATE: 29-MAY-1992  
; CLASSIFICATION: 514  
; CLASSIFICATION: A 61 K 37/64  
; CLASSIFICATION: A 61 K 47/48  
; PRIOR APPLICATION NUMBER: PCT/EP90/01998  
; FILING DATE: 22-NOV-1990  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 59 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
US-07-859-453E-9

Query Match 78.9%; Score 30; DB 1; Length 59;  
Best Local Similarity 86.8%; Pred. No. 0.0012;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggctctagagcgcaaaaaatcaatgcgttactggcgca 38

Db 16 GGCTCTAAGGCCAGAAAAAACCACTGCGTACTGGCGA 53

## RESULT 3

US-08-262-384A-5/c  
; Sequence 5, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 128 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-262-384A-5

Query Match 78.9%; Score 30; DB 1; Length 128;  
Best Local Similarity 86.8%; Pred. No. 0.0013;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggctctagagcgcaaaaaatcaatgcgttactggcgca 38

Db 116 GGCTCTGACGGCGAAGAAAAAACCACTGCGTACTGGCGA 79

## RESULT 4

US-08-262-384A-7  
; Sequence 7, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: 530  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 235 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

## US-08-262-384A-7

Query Match 78.9%; Score 30; DB 1; Length 235;  
Best Local Similarity 86.8%; Pred. No. 0.0014;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggctctagagcgcaaaaaatcaatcgcttactgcca 38  
||||| ||||||||| ||||||||| ||||||||| |||  
Db 124 GGCTCTGACGGCGAAGAACACGTCGCTTACTGCGCA 161

## RESULT 5

US-08-262-384A-8/c  
; Sequence 8, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820  
; FILING DATE: 18-JUN-1992  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 726 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-262-384A-9

Query Match 78.9%; Score 30; DB 1; Length 726;  
Best Local Similarity 86.8%; Pred. No. 0.0017; 5; Indels 0; Gaps 0;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggctctagagcgcaaaaaatcaatcgcttactgcca 38  
||||| ||||||||| ||||||||| ||||||||| |||  
Db 616 GGCTCTGACGGCGAAGAACACGTCGCTTACTGCGCA 653

## RESULT 7

US-08-262-384A-10/c  
; Sequence 10, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.0  
; SOFTWARE: WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/262,384A  
; FILING DATE: 20-JUN-1994  
; CLASSIFICATION: C 12 N 15/62  
; CLASSIFICATION: C 12 N 15/31  
; CLASSIFICATION: C 07 K 7/10  
; CLASSIFICATION: C 12 P 21/02  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP90/02084  
; FILING DATE: 04-DEC-1990  
; APPLICATION NUMBER: US 07861820

Query Match 78.9%; Score 30; DB 1; Length 235;  
Best Local Similarity 86.8%; Pred. No. 0.0014;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggctctagagcgcaaaaaatcaatcgcttactgcca 38  
||||| ||||||||| ||||||||| ||||||||| |||  
Db 116 GGCTCTGACGGCGAAGAACACGTCGCTTACTGCGCA 79

## RESULT 6

US-08-262-384A-9  
; Sequence 9, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff

;; FILING DATE: 18-JUN-1992  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 726 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
US-08-262-384A-10

Query Match 78.9%; Score 30; DB 1; Length 726;  
Best Local Similarity 86.8%; Pred. No. 0.0017; 5; Indels 0; Gaps 0;  
Matches 33; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgaaaaataatcggttactggcgga 38  
||||| ||||||| || ||||||| |||||  
Db 111 GGCTCTGACGCGGAAAAAACCAAGTGGTACTGGCGA 74

RESULT 8  
US-07-859-453E-11  
; Sequence 11, Application US/07859453E  
; Patent No. 5663141  
; GENERAL INFORMATION:  
; APPLICANT: Kurfuerst, Manfred  
; APPLICANT: Ruebsamen, Klaus  
; APPLICANT: Schmied, Bernhard  
; APPLICANT: Koerwer, Wolfgang  
; APPLICANT: Schweden, Juergen  
; APPLICANT: Hoefken, Hans Wolfgang  
; TITLE OF INVENTION: Hrudin/polyalkylene glycol  
; TITLE OF INVENTION: conjugates  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinlauf  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS version 7.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/859,453E  
FILING DATE: 29-MAY-1992  
CLASSIFICATION: 514  
CLASSIFICATION: A 61 K 37/64  
CLASSIFICATION: A 61 K 47/48  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/01998  
FILING DATE: 22-NOV-1990  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 59 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
US-07-859-453E-11

Query Match 74.7%; Score 28.4; DB 1; Length 59;  
Best Local Similarity 84.2%; Pred. No. 0.0051;  
Matches 32; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgaaaaataatcggttactggcgga 38  
||||| ||||||| || ||||||| |||||  
Db 16 GGCTCTAAGCGCAACGTAACCAAGTGGTACTGGCGA 53

RESULT 9

US-08-044-506B-27  
; Sequence 27, Application US/08044506B  
; Patent No. 5552299  
; GENERAL INFORMATION:  
; APPLICANT: OTT, Istvan  
; APPLICANT: KLUPP, Tibor  
; APPLICANT: MOLNAR, Istvan  
; APPLICANT: PATTHY, Andras  
; APPLICANT: BARTA, Istvan  
; APPLICANT: BARKO nee TOTH, Zsuzsa  
; APPLICANT: AMBRUS, Gabor  
; APPLICANT: SALAT, Janos  
; APPLICANT: TEGDES, Aniko  
; APPLICANT: MORAVCSIK, Imre  
; TITLE OF INVENTION: Plasmids and Process for Producing  
; TITLE OF INVENTION: Recombinant Desulphatohirudin HV-1 Peptides  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinlauf  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
COMPUTER: IBM AT-compatible, 80286 processor  
OPERATING SYSTEM: MS-DOS version 5.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/044,506B  
FILING DATE: 09-APR-1993  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 195 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-044-506B-27

Query Match 73.7%; Score 28; DB 1; Length 195;  
Best Local Similarity 81.6%; Pred. No. 0.0088;  
Matches 31; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgaaaaataatcggttactggcgga 38  
||||| ||||||| || ||||||| |||||  
Db 91 GGTTCTACGGTGAAAAAATCAATGTCTACTGGCGA 128

RESULT 10  
US-08-044-506B-22  
; Sequence 22, Application US/08044506B  
; Patent No. 5552299  
; GENERAL INFORMATION:  
; APPLICANT: OTT, Istvan  
; APPLICANT: KLUPP, Tibor  
; APPLICANT: MOLNAR, Istvan  
; APPLICANT: PATTHY, Andras  
; APPLICANT: BARTA, Istvan  
; APPLICANT: BARKO nee TOTH, Zsuzsa  
; APPLICANT: AMBRUS, Gabor  
; APPLICANT: SALAT, Janos  
; APPLICANT: TEGDES, Aniko  
; APPLICANT: MORAVCSIK, Imre  
; TITLE OF INVENTION: Plasmids and Process for Producing  
; TITLE OF INVENTION: Recombinant Desulphatohirudin HV-1 Peptides  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinlauf  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington



Search completed: July 15, 2002, 22:31:52  
Job time: 18478 sec

FILING DATE: 435  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/672,205  
 FILING DATE: 19-MAR-1991  
 APPLICATION NUMBER: GB 9006400.7  
 FILING DATE: 22-MAR-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Villamizar, Joann  
 REGISTRATION NUMBER: 30,598  
 REFERENCE/DOCKET NUMBER: 4-17994/A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (914)785-7121  
 TELEFAX: (914)347-5769  
 INFORMATION FOR SPO ID NO: 3:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (recombinant)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Hybrid gene: MSP signal (base pair 1 to 81),/
INDIVIDUAL ISOLATE: desulfatohirudin (base pair 82 to 279)
IMMEDIATE SOURCE:
CLONE: pUCRS/pML310
FEATURE:
NAME/KEY: CDS
LOCATION: 1..276
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 82..276
US-08-186-222-3

```

Query Match	66.3%	Score 25.2;	DB 1;	Length 279;
Best Local Similarity	78.9%	Pred. No. 0.12;		
Matches 30;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0;
Qy	1	ggctctagagcgcgaataaatcaatgcgtactggcgca	38	
Db	172	ggttctgacgggtgaaaaaaacacagctgttaccgcgca	209	

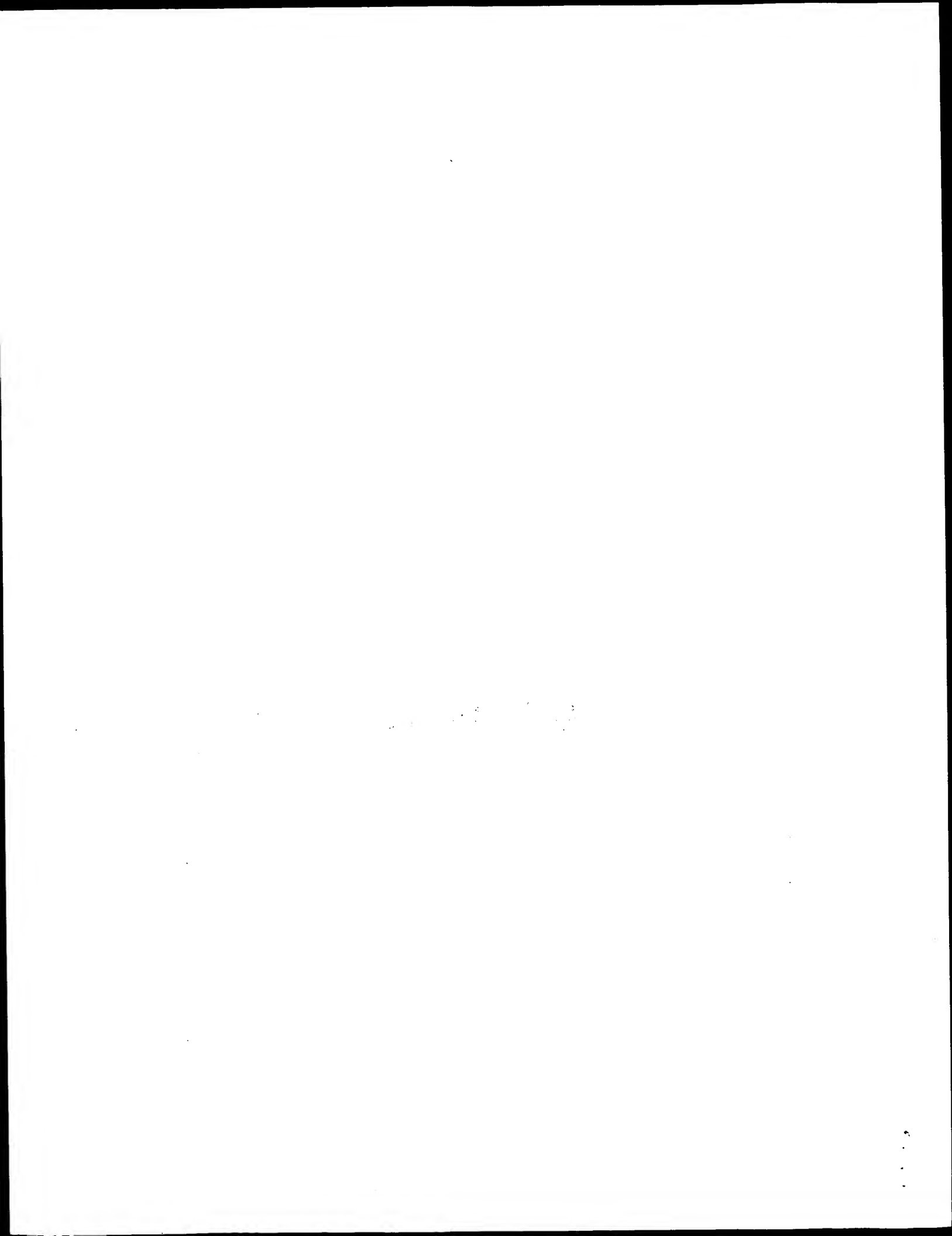
```

RESULT 15
5166318-16
; Patent No. 5166318
; APPLICANT: FURUTANI, YOSHIO; HONJO, MASARU; NAKAYAMA, AKIRA;
; KAWAMURA, KOICHI; ANDO, KAZUNORI; HORI, MICHIO; FUKAZAWA, KEIKO
; KAWAMURA, KOICHI; ANDO, KAZUNORI; HORI, MICHIO; FUKAZAWA, KEIKO
; TITLE OF INVENTION: POLYPEPTIDE HAVING THROMBIN INHIBITING
; ACTIVITY
; NUMBER OF SEQUENCES: 22
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/531,403
; FILING DATE: 04-JUN-1990
; SEQ ID NO:16:
; LENGTH: 526
5166318-16

```

Query Match' 66.3%; Score 25.2; DB 6; Length 526;  
Best Local Similarity 78.9%; Pred. No. 0.13;  
Matches 30; Conservative 8; Mismatches 0; Gaps 0;  
Ov 1 gactctagaggccgaaataatcaatgcgttactgcga 38







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:49:15 ; Search time 6165.88 Seconds  
(without alignments)  
83.181 Million cell updates/sec

Title: US-10-053-641-7  
Perfect score: 38  
Sequence: 1 ggctctagagcgcaaaaaaatcaatcggttactgcgca 38

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_gss:\*

13: em\_gss\_hum:\*

14: em\_gss\_inv:\*

15: em\_gss\_pln:\*

16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	60.5	477	9	AW330364
2	21.6	56.8	473	9	AW634909
3	21.6	56.8	938	12	CNS03ACU
4	21.4	56.3	311	12	AL235047 Tetraodon
5	21.4	56.3	451	12	AQ005962 CIT-HSP-2
6	21.4	56.3	503	9	BH048008 RPI-24-3
7	21.4	56.3	682	10	AW330322
8	21.4	56.3	857	12	C96522
9	21.2	55.8	386	10	BH439433
10	21.2	55.8	476	9	BI972489
11	21.2	55.8	559	9	AI900286
12	21.2	55.8	678	12	AW831943
13	21.2	55.3	376	10	A2569434
14	21	55.3	600	10	C64633
15	20.8	54.7	461	12	BG808735
16	20.8	54.7	645	12	AQ119701
17	20.8	54.7	656	12	AQ235385
					AG176441 Pan trogl

C	18	20.6	54.2	313	12	AQ103411
	19	20.6	54.2	390	9	AA921897
	20	20.6	54.2	428	9	BB638076
	21	20.6	54.2	708	9	BB178030
C	22	20.6	54.2	725	9	BB096074
	23	20.6	54.2	727	9	BB363943
	24	20.6	54.2	797	12	AQ748677
	25	20.6	54.2	1020	12	CNS05JFK
	26	20.4	53.7	229	10	BF891361
C	27	20.4	53.7	344	9	AA215143
C	28	20.4	53.7	404	9	AA706088
	29	20.4	53.7	426	9	AW653178
C	30	20.4	53.7	427	10	R01409
	31	20.4	53.7	449	12	AZ847809
	32	20.4	53.7	495	9	AI722825
C	33	20.4	53.7	550	9	AA034037
	34	20.4	53.7	600	10	BI985526
	35	20.4	53.7	600	10	BI985526
	36	20.4	53.7	627	9	BI988827
	37	20.4	53.7	686	9	BB642983
	38	20.4	53.7	719	12	AG112884
C	39	20.4	53.7	728	9	AI855745
C	40	20.4	53.7	806	12	BH357948
	41	20.4	53.7	816	10	BW016367
C	42	20.4	53.7	920	9	AA826328
	43	20.4	53.7	1047	10	BI553059
C	44	20.2	53.2	241	9	AV250045
C	45	20.2	53.2	389	9	AW488556

## ALIGNMENTS

## RESULT 1

AW330364  
LOCUS  
DEFINITION  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477

AW330364  
TENU5110 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
cruzi cDNA clone 44h15 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Trypanosoma cruzi.  
Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae;  
Trypanosoma; Schizotrypanum.  
REFERENCE  
1 (bases 1 to 477)  
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.  
Gene survey of the pathogenic protozoan Trypanosoma cruzi  
Genome Res. 10 (8), 1103-1107 (2000)  
20414748  
Contact: Aslund L  
Department of Medical Genetics  
Uppsala University  
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
Tel: 46 18 471 45 85  
Fax: 46 18 52 68 49  
Email: lena.aslund@medgen.uu.se  
Seq primer: 17 primer  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 477



## RESULT 4

AQ005962/c

LOCUS

DEFINITION

CIT-HSP-2292120.TR CIT-HSP Homo sapiens genomic clone 2292120, DNA

sequence.

ACCESSION

AQ005962

VERSION

AQ005962.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 311)

ADAMS,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,

Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,

Simon,M., and Venter,J.C.

Use of a random BAC End Sequence Database for Sequence-Ready Map

Building (1998)

JOURNAL

Unpublished (1998)

COMMENT

Other\_GSSs: CIT-HSP-2292120.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com).

BAC

end search page:

http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html

Seq primer: M13 Reverse

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..311

/organism="Homo sapiens"

/db\_xref="GDB:7151471"

/db\_xref="taxon:9606"

/clone="2292120"

/clone\_lib="CIT-HSP"

/sex="Male"

/cell\_type="Sperm"

/note="Vector: pBelOBAcl1; Site\_1: HindIII; Site\_2:

HindIII"

71 a 74 c 67 g 99 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

25; Conservative

0; Mismatches

6; Indels

0; Gaps

0;

QY

1

ggctctgagggcgagaaataatcgcttta

31

Db

43

GTCTCTAAGGGGAGAGAAATCAGTGGTTA

13

RESULT 5

BH048008

LOCUS

DEFINITION

RPCI-24-330011.TVB RPCI-24 Mus musculus genomic clone

RPCI-24-330011, DNA sequence.

ACCESSION

BH048008

VERSION

BH048008.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 451)

Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,

Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,

Russell,D., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-24

TITLE

source

Location/Qualifiers

1..503

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-24-330011"

/clone\_lib="RPCI-24"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;

RPCI-24 Mouse BAC library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

170 a 55 c 82 g 144 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

25; Conservative

0; Mismatches

6; Indels

0; Gaps

0;

QY

5

ctagaggcgagaaataatcgcttactgg

35

Db

136

CTAGAGTCAAAAAAATTCATGCAATATGG

166

RESULT 6

AW330322

LOCUS

DEFINITION

TENU5068 T.cruzi epimastigote normalized cDNA Library Trypanosoma

cruzi cDNA clone 44e7 5', mRNA sequence.

ACCESSION

AW330322

VERSION

AW330322.1

KEYWORDS

EST.

SOURCE

Trypanosoma cruzi.

ORGANISM

Trypanosoma cruzi

REFERENCE

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.

1 (bases 1 to 503)

Forcel,B.M., Tran,A.-N., Tammi,M., Nyarady,2., Rydaker,M., Urmenyi

,T.P., Rondinelli,E., Pettersson,U., Andersson,B. and Aslund,L.

Gene survey of the pathogenic protozoan Trypanosoma cruzi

Genome Res. 10 (8), 1103-1107 (2000)

2041478

COMMENT

Contact: Aslund L

Department of Medical Genetics

Uppsala University

Biomedical Center, Box 589, S-751 23 Uppsala, Sweden

Tel: 46 18 471 45 85

Fax: 46 18 52 68 49

Email: lena.aslund@medgen.uu.se

Seq primer: T7 primer

High quality sequence stop: 503.

FEATURES

source

Location/Qualifiers

1..503

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-24-330011"

/clone\_lib="RPCI-24"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;

RPCI-24 Mouse BAC library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

170 a 55 c 82 g 144 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

25; Conservative

0; Mismatches

6; Indels

0; Gaps

0;

QY

5

ctagaggcgagaaataatcgcttactgg

35

Db

136

CTAGAGTCAAAAAAATTCATGCAATATGG

166

RESULT 6

AW330322

LOCUS

DEFINITION

TENU5068 T.cruzi epimastigote normalized cDNA Library Trypanosoma

cruzi cDNA clone 44e7 5', mRNA sequence.

ACCESSION

AW330322

VERSION

AW330322.1

KEYWORDS

EST.

SOURCE

Trypanosoma cruzi.

ORGANISM

Trypanosoma cruzi

REFERENCE

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.

1 (bases 1 to 503)

Forcel,B.M., Tran,A.-N., Tammi,M., Nyarady,2., Rydaker,M., Urmenyi

,T.P., Rondinelli,E., Pettersson,U., Andersson,B. and Aslund,L.

Gene survey of the pathogenic protozoan Trypanosoma cruzi

Genome Res. 10 (8), 1103-1107 (2000)

2041478

COMMENT

Contact: Aslund L

Department of Medical Genetics

Uppsala University

Biomedical Center, Box 589, S-751 23 Uppsala, Sweden

Tel: 46 18 471 45 85

Fax: 46 18 52 68 49

Email: lena.aslund@medgen.uu.se

Seq primer: T7 primer

High quality sequence stop: 503.

FEATURES

source

Location/Qualifiers

1..503

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-24-330011"

/clone\_lib="RPCI-24"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;

RPCI-24 Mouse BAC library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

170 a 55 c 82 g 144 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

25; Conservative

0; Mismatches

6; Indels

0; Gaps

0;

QY

5

ctagaggcgagaaataatcgcttactgg

35

Db

136

CTAGAGTCAAAAAAATTCATGCAATATGG

166

RESULT 6

AW330322

LOCUS

DEFINITION

TENU5068 T.cruzi epimastigote normalized cDNA Library Trypanosoma

cruzi cDNA clone 44e7 5', mRNA sequence.

ACCESSION

AW330322

VERSION

AW330322.1

KEYWORDS

EST.

SOURCE

Trypanosoma cruzi.

ORGANISM

Trypanosoma cruzi

REFERENCE

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.

1 (bases 1 to 503)

Forcel,B.M., Tran,A.-N., Tammi,M., Nyarady,2., Rydaker,M., Urmenyi

,T.P., Rondinelli,E., Pettersson,U., Andersson,B. and Aslund,L.

Gene survey of the pathogenic protozoan Trypanosoma cruzi

Genome Res. 10 (8), 1103-1107 (2000)

2041478

COMMENT

Contact: Aslund L

Department of Medical Genetics

Uppsala University

Biomedical Center, Box 589, S-751 23 Uppsala, Sweden

Tel: 46 18 471 45 85

Fax: 46 18 52 68 49

Email: lena.aslund@medgen.uu.se

Seq primer: T7 primer

High quality sequence stop: 503.

FEATURES

source

Location/Qualifiers

1..503

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="R

```

/organism="trypanosoma cruzi"
/strain="Cl-Brenner"
/db_xref="taxon:5693"
/clone="44e7"
/cell_lib="r.cruzi epimastigote normalized cDNA Library"
/note="cDNA library constructed with oligo dt primed
epimastigote mRNA and cloned in pT7318D phagemid with
modified polylinker (Pharmacia)"
BASE COUNT      148 a 102 c 108 g 145 t
ORIGIN

Query Match      56.3%; Score 21.4; DB 9; Length 503;
Best Local Similarity 80.6%; Pred. NO. 1.2e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcgctta 31
    ||||| ||| ||||| ||| ||| |||
Db 440 GTCTCTAAGGGGAAAAAATGACTGCATTA 470

RESULT 7
LOCUS C96522 682 bp mRNA linear EST 27-JUL-1999
DEFINITION C96522 Marchantia polymorpha immature sex organ Marchantia
ACCESSION polymorpha cDNA clone F01W094, mRNA sequence.
VERSION C96522
KEYWORDS EST.
SOURCE C96522.1 GI:4719400
ORGANISM Marchantia polymorpha
Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Marchantiophyta; Marchantiopsida; Marchantiales;
Marchantiaceae; Marchantiaceae; Marchantia.
REFERENCE 1 (bases 1 to 682)
AUTHORS Nagai,J., Yamato,K.T., Sakaida,M., Yoda,H., Fukuzawa,H. and Ohyama
,K.
TITLE Expressed sequence tags from immature female sexual organ of a
liverwort, Marchantia polymorpha
JOURNAL DNA Res. 6 (1), 1-11 (1999)
MEDLINE 99246055
COMMENT Contact: Nagai J
Graduate School of Agriculture
Kyoto University, Division of Applied Life Sciences
Kitashirakawa-oiwake-cho, Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-6389
Fax: 81-75-753-6127
Email: kohnyama@kais.kyoto-u.ac.jp.
Location/Qualifiers
1..682
/organism="Marchantia polymorpha"
/db_xref="taxon:3197"
/clone="F01W094"
/clone_lib="Marchantia polymorpha immature sex organ"
/sex="female"
/tissue_type="immature sex organ"
BASE COUNT      177 a 155 c 177 g 173 t
ORIGIN

Query Match      56.3%; Score 21.4; DB 10; Length 682;
Best Local Similarity 80.6%; Pred. NO. 1.2e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ggctctagagcgcaaaaaaatcaatcgctta 31
    ||||| ||| ||||| ||| ||| |||
Db 76 GGCTCTCAAGGAGAAAAAATCAAGTGTTA 106

RESULT 8
LOCUS BH439433 857 bp DNA linear GSS 12-DEC-2001

```

```

DEFINITION BOGUF47TR BOGU Brassica oleracea genomic clone BOGUF47, DNA
sequence.
ACCESSION BH439433
VERSION BH439433.1 GI:17625147
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 857)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOGUF47TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1..857
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGUF47"
/clone_lib="BOGU"
/note="vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT      216 a 185 c 201 g 255 t
ORIGIN

Query Match      56.3%; Score 21.4; DB 12; Length 857;
Best Local Similarity 80.6%; Pred. NO. 1.1e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 tctagagcgcaaaaaaatcaatcgcttactg 34
    ||||| ||||| ||| ||| ||| |||
Db 742 TCTAGAGCGCAACAAAGTTCAAGGCGACACTG 772

RESULT 9
LOCUS BI972489 386 bp mRNA linear EST 30-NOV-2001
DEFINITION sai78c03.v1 Gm-cl065 Glycine max cDNA clone GENOME SYSTEMS CLONE
ACCESSION ID: Gm-cl065-7014 5', mRNA sequence.
VERSION BI972489
KEYWORDS EST.
SOURCE BI972489.1 GI:16346894
ORGANISM soybean.
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 386)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

```

Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130  
South Memorial Parkway Hunttsville, AL 35801 For further information  
call: (800)-533-4363 or contact via email: ccu@resgen.com  
High quality sequence stop: 376.

## FEATURES

source

Location/Qualifiers  
1. 386  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1065-7014"  
/clone\_lib="Gm-c1065"  
/tissue\_type="germinating shoots"  
/lab\_host="DH10B"

/note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2: XhoI; The cDNA library was constructed from mRNA isolated from germinating shoots of the cultivar Williams. The seeds were allowed to germinate for 24 hours prior to being cold stressed for 2 days at 4C. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT 67 a 142 c 54 g 123 t

## Query Match

Best Local Similarity 55.8%; Score 21.2; DB 10; Length 386;

Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ctctagaggcgaaaaaatcaatcg 28

||||||| |||||||||

Db 21 CTCTAGAGGTTGAAAAATCAATCG 46

## RESULT 10

AI900286

LOCUS AI900286

DEFINITION sc03e01.y1 Gm-c1012 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:

Gm-c1012-1105 5', mRNA sequence.

ACCESSION AI900286

VERSION AI900286.1 GI:5606188

KEYWORDS EST

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

## REFERENCE

AUTHORS

1 (bases 1 to 476)

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130

## FEATURES

source

Location/Qualifiers  
1. 476  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1012-1105"  
/clone\_lib="Gm-c1012"  
/tissue\_type="Apical shoot tips, 9-10 day old etiolated seedlings"  
/lab\_host="XL10-Gold"

/note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2: XhoI; This cDNA library was constructed from mRNA isolated from the apical shoots of 9 to 10 day old etiolated seedlings. The shoot tips including any emerged leaves were harvested for mRNA isolation. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 90 a 154 c 79 g 152 t 1 others

## Query Match

Best Local Similarity 55.8%; Score 21.2; DB 9; Length 476;

Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ctctagaggcgaaaaaatcaatcg 28

||||||| |||||||||

Db 22 CTCTAGAGGTTGAAAAATCAATCG 47

## RESULT 11

AW831943

LOCUS AW831943

DEFINITION sm18a11.y1 Gm-c1027 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:

Gm-c1027-9237 5', mRNA sequence.

ACCESSION AW831943

VERSION AW831943.1 GI:7925917

KEYWORDS EST

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

## REFERENCE

AUTHORS

1 (bases 1 to 559)

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information  
call: (800)-533-4363 or contact via email: ccuresgen.com  
Insert Length: 1500 Std Error: 0.00  
High quality sequence stop: 443.

## FEATURES

source

```

1. .559
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl027-9237"
/clone_lib="Gm-cl027"
/tissue_type="cotyledons of 3- and 7-day-old Williams
seedlings"
/lab_host="DH10B"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; This cDNA library was constructed from mRNA isolated
from cotyledons of 3- and 7-day-old Williams seedlings
which were propagated on paper towels with distilled
water. The cotyledons were flash-frozen in liquid
nitrogen, then lyophilized for 72 hours. Unequal amounts
of mRNA was used for cDNA synthesis. Stratagene's cDNA
Synthesis Kit (catalog number 200401) was used to
synthesize the cDNA. First-strand synthesis was
performed with 5-methyl dCTP, hence the ligated cDNA was
hemimethylated. A modification of Stratagene's
first-strand synthesis primer was used. An anchor
nucleotide (V-A, C, or G) was added to the 3' end of the
primer [GAGAGAGAGAGAGAGAGAGAGAGAG(T)18] to anchor
the primer at the 5' end of the poly(A) tract. After
second-strand synthesis, the cDNA ends were filled in
with cloned Pfu DNA, ligated to EcoRI adapters and
subsequently phosphorylated. The XhoI site within the
first-strand synthesis primer was then restricted by
digestion with XhoI; all XhoI sites in the cDNA would be
protected by their hemimethylated status. The cDNA
constructs were size-fractionated with a 500 bp cutoff,
using GibcoBRL life technologies' cDNA size fractionation
column. The column eluent was then ligated into
Stratagene's pBluescript(tm) II XR predigested vector
(pBluescript II SK+) that has been digested with EcoRI
and XhoI, and phosphorylated by Stratagene). 97% of the
white and blue colonies appear to contain recombinant
plasmids with cDNA inserts, based on size (n=30). This
library was constructed by Dr. Paul Keim and Dr. Virginia
Coryell."

```

```

BASE COUNT      115 a 176 c 90 g 177 t 1 others
ORIGIN

Query Match      55.8%; Score 21.2; DB 9; Length 559;
Best Local Similarity 88.5%; Pred. No. 1.4e+03;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 3 ctctagagcgcaaaaatcaatgcg 28
|||||  |||||  |||||  |||||  |||||  |||||
DB 9 CTCTAGAGTTGAAAATCAATGCG 34

```

```

RESULT 12
AZ569434 AZ569434 678 bp DNA linear GSS 15-MAY-2001
LOCUS 260PVG03 Pv MBN #30 Plasmodium vivax genomic 3', DNA sequence.
DEFINITION
ACCESSION AZ569434.1 GI:13979525
VERSION
KEYWORDS GSS.
SOURCE malaria parasite P. vivax.
ORGANISM Plasmodium vivax
Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 678)
AUTHORS Carlton,J.M.-R. and Dame,J.B.
TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL Parasitol. Today 16 (10), 409 (2000)
COMMENT Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine

```

University of Florida  
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA  
Tel: 352 392 4700  
Fax: 352 392 9704  
Email: damej@mail.vetmed.ufl.edu  
Seq primer: M13(-20) forward  
Class: shotgun.

## FEATURES

source

```

1. .678
/organism="plasmodium vivax"
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
497-598)"
/db_xref="taxon:5855"
/clone_lib="Pv MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Saimiri boliviensis"
/notes="Vector: pBluescript SK(+) vector DNA, phagemid
excised from lambda ZAP; Site 1: EcoR V; Site 2: EcoR V;
Host leukocytes were extracted from P. vivax infected blood
using the following methods: first, infected blood
was activated by the addition of 0.5 ml of ADP (40mg/ml)
per 10 ml blood. Then blood was passed over a column of
acid washed 0.1 mm glass beads, then through a plasmidpur
filter, followed by passage through a column of pre-wet
Whatman CFF powder (1:2 ratio volume of blood to CFF),
and finally centrifuged through a 50% Percoll density
cushion. Purified DNA was digested with mung bean nuclease
in the presence of 44% formamide at 500c as described
(Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). Digested DNA was
blunt-ended using T4 DNA polymerase and size fractionated
over a Sepharose CL-2B column. Fractions in the size range
500bp-4kb were ligated into the Eco RV site of pBluescript
SK(+), and E. coli XL-10 Gold transformed with the
ligation mixture."

```

```

BASE COUNT      157 a 188 c 156 g 177 t
ORIGIN

```

```

Query Match      55.8%; Score 21.2; DB 12; Length 678;
Best Local Similarity 76.5%; Pred. No. 1.4e+03;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 1 ggcctagagcgcaaaaatcaatgcgttactg 34
|||||  |||||  |||||  |||||  |||||  |||||
DB 416 GGGTAAAGGCGAAAAGAGACGCGTCTCTG 449

```

## RESULT 13

```

C64633 C64633 376 bp mRNA linear EST 22-SEP-1997
LOCUS C64633 Yujii Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION
ACCESSION C64633
VERSION C64633.1 GI:2423338
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

```

```

REFERENCE 1 (bases 1 to 376)
AUTHORS kohara,Y., Motoshashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yujii Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

```

```

source
1. .376
/organism="Caenorhabditis elegans"
/strain="CBL489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk357d3"
/clone_lib="Yuj1 Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT      143 a      93 g      61 t      1 others
ORIGIN

Query Match      55.3%; Score 21; DB 10; Length 376;
Best Local Similarity 73.0%; Pred. No. 1.6e+03;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 gctctagaggcgagaaatacaatcgcttactgagcga 38
||||| ||| ||||| ||||| ||||| ||||| |||||
Db 319 GATCAACAGCTCAAAAAATAAATGCGTGTCTCGAGA 355

RESULT 14
BG808735
LOCUS
DEFINITION
2121-62 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
mRNA sequence.
ACCESSION
BG808735
VERSION
BG808735.1 GI:17955718
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
White, R.A., Beremand, P.D., Thomas, F.L., Gan, L. and Klein, W.H.
Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
Nucleic Acids Res. 29 (24), 4983-4993 (2001)
Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.

FEATURES
source
1. .600
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"

BASE COUNT      162 a      118 c      123 g      189 t      8 others
ORIGIN

Query Match      55.3%; Score 21; DB 10; Length 600;
Best Local Similarity 71.1%; Pred. No. 1.6e+03;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ggctctagaggcgagaaatacaatcgcttactgagcga 38
||||| ||| ||||| ||||| ||||| ||||| |||||
Db 450 GGCTCTNTGGGAAAAAATCCATCGGTGACAGAAGA 487

RESULT 15
AQ119701/c
LOCUS
DEFINITION
HS_2171_B2_H04_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-2171 Col-8 Row=P, DNA sequence.
ACCESSION
AQ119701
VERSION
AQ119701.1 GI:3494468

KEYWORDS
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 461)
AUTHORS
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
TITLE
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE
99380589
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2171 row: P column: 8
Class: BAC ends
High quality sequence stop: 461.
Location/Qualifiers
1. .461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2171 Col=8 Row=P"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="organ: sperm; Vector: pBel8BAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT      102 a      76 c      97 g      183 t      3 others
ORIGIN

Query Match      54.7%; Score 20.8; DB 12; Length 461;
Best Local Similarity 75.8%; Pred. No. 1.9e+03;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 6 tagaggcgagaaaaatacaatcgcttactgagcga 38
||||| ||| ||||| ||||| ||||| ||||| |||||
Db 56 TTGATGCGNAAAAAACATCAATAAGATACTGGCAA 24

Search completed: July 15, 2002, 21:49:22
Job time: 18913 sec

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:29:30 ; Search time 2368.24 Seconds  
(without alignments)  
282.762 Million cell updates/sec

Title: US-10-053-641-8  
Perfect score: 32  
Sequence: 1 catcgccgcctattattgcaggattctt 32

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	65.6	34	6	182199	I82199	Sequence 13 from patent US 5712114.
2	65.0	76	6	A31395	I82199	Sequence 13 from patent US 5712114.
3	65.0	76	6	A31398	I82199	Sequence 13 from patent US 5712114.
4	65.0	128	6	I41185	I82199	Sequence 13 from patent US 5712114.
5	65.0	235	6	I41188	I82199	Sequence 13 from patent US 5712114.
6	62.5	197369	9	CNSOIRH6	I82199	Sequence 13 from patent US 5712114.
7	60.6	68	6	A31391	I82199	Sequence 13 from patent US 5712114.
8	60.6	68	6	A31394	I82199	Sequence 13 from patent US 5712114.
9	60.6	68	6	A31397	I82199	Sequence 13 from patent US 5712114.
10	60.6	68	6	I63550	I82199	Sequence 13 from patent US 5712114.
11	60.6	68	6	I63552	I82199	Sequence 13 from patent US 5712114.
12	60.6	68	6	I63554	I82199	Sequence 13 from patent US 5712114.
13	60.6	68	6	A31392	I82199	Sequence 13 from patent US 5712114.
14	60.6	106	6	I41184	I82199	Sequence 13 from patent US 5712114.
15	60.6	226	12	SYNHIR	I82199	Sequence 13 from patent US 5712114.
16	60.6	235	6	I41187	I82199	Sequence 13 from patent US 5712114.
17	60.6	854	33	AC052752	I82199	Sequence 13 from patent US 5712114.
18	60.6	952	33	AC083648	I82199	Sequence 13 from patent US 5712114.
19	60.6	1000	33	AC059290	I82199	Sequence 13 from patent US 5712114.
20	60.6	1086	10	MM0277752	I82199	Sequence 13 from patent US 5712114.
21	60.6	51923	9	AL356983	I82199	Sequence 13 from patent US 5712114.
22	60.6	215550	2	AL671299	I82199	Sequence 13 from patent US 5712114.
23	60.6	215734	2	AC091784	I82199	Sequence 13 from patent US 5712114.
24	60.0	1778	3	AB018047	I82199	Sequence 13 from patent US 5712114.
25	60.0	30848	3	AC004307	I82199	Sequence 13 from patent US 5712114.
26	60.0	46383	2	AC020300	I82199	Sequence 13 from patent US 5712114.
27	60.0	74717	3	AC004563	I82199	Sequence 13 from patent US 5712114.
28	60.0	83693	8	AT133J8	I82199	Sequence 13 from patent US 5712114.
29	60.0	140310	2	AC016009	I82199	Sequence 13 from patent US 5712114.
30	60.0	156617	2	AC099151	I82199	Sequence 13 from patent US 5712114.
31	60.0	156799	2	AC099151	I82199	Sequence 13 from patent US 5712114.
32	60.0	179078	2	AC079608	I82199	Sequence 13 from patent US 5712114.
33	60.0	197278	9	AC007367	I82199	Sequence 13 from patent US 5712114.
34	60.0	199749	8	ATCHRIV68	I82199	Sequence 13 from patent US 5712114.
35	60.0	259914	3	AE003795	I82199	Sequence 13 from patent US 5712114.
36	60.0	304290	1	CNSPAX05	I82199	Sequence 13 from patent US 5712114.
37	60.0	349980	6	AX041920	I82199	Sequence 13 from patent US 5712114.
38	60.0	349980	6	AX041921	I82199	Sequence 13 from patent US 5712114.
39	59.4	400	11	G31622	I82199	Sequence 13 from patent US 5712114.
40	59.4	400	11	G34180	I82199	Sequence 13 from patent US 5712114.
41	59.4	28763	2	AC103135	I82199	Sequence 13 from patent US 5712114.
42	59.4	42458	2	AC103285	I82199	Sequence 13 from patent US 5712114.
43	59.4	63806	9	AL450340	I82199	Sequence 13 from patent US 5712114.
44	59.4	144404	8	AP003210	I82199	Sequence 13 from patent US 5712114.
45	59.4	145410	9	AC023281	I82199	Sequence 13 from patent US 5712114.

RESULT	1	21	34	6	182199	34 bp	DNA	linear	PAT 10-JUN-1998
LOCUS	I82199	Sequence 13 from patent US 5712114.							
DEFINITION	I82199	Sequence 13 from patent US 5712114.							
ACCESSION	I82199	Sequence 13 from patent US 5712114.							
VERSION	I82199.1	GI:3210496							
KEYWORDS									
SOURCE	Unknown.								
ORGANISM	Unknown.								
REFERENCE	1 (bases 1 to 34)								
AUTHORS	Mankovich, J.A., Hammill, L. and Ferenz, C.R.								
TITLE	Compositions for expression of proteins in host cells using a								
JOURNAL	Preprocollagen signal sequence								
FEATURES	Patent: US 5712114-A 13 27-JAN-1998;								
Source	Location/Qualifiers								
	1..34								
	/organism="unknown"								
BASE COUNT	5 a	8 c	11 g	10 t					
ORIGIN									

Query Match 65.6% Score 21; DB 6; Length 34;

[illegible]



```

REFERENCE 1 (bases 1 to 68)
AUTHORS
TITLE HIRUDINE POLYALKYLENE GLYCOL CONJUGATES
JOURNAL Patent: WO 9108229-A 14 13-JUN-1991;
FEATURES Location/Qualifiers
source
BASE COUNT 21 a 21 c 17 g 9 t
ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 68;
Best Local Similarity 95.2%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32
||||| |||||||
DB 68 CCTATTACTGCAGGTATTCTT 48

RESULT 10
I63550/c
LOCUS I63550 68 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 13 from patent US 5663141.
ACCESSION I63550
VERSION I63550.1 GI:2481123
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68)
AUTHORS Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
Hoeffken,H.Wolfgang.
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL Patent: US 5663141-A 13 02-SEP-1997;
FEATURES Location/Qualifiers
source
BASE COUNT 23 a 20 c 16 g 9 t
ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 68;
Best Local Similarity 95.2%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32
||||| |||||||
DB 68 CCTATTACTGCAGGTATTCTT 48

RESULT 11
I63552/c
LOCUS I63552 68 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 15 from patent US 5663141.
ACCESSION I63552
VERSION I63552.1 GI:2481125
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68)
AUTHORS Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
Hoeffken,H.Wolfgang.
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL Patent: US 5663141-A 15 02-SEP-1997;
FEATURES Location/Qualifiers
source
BASE COUNT 20 a 21 c 17 g 10 t
ORIGIN

REFERENCE 1 (bases 1 to 68)
AUTHORS Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
Hoeffken,H.Wolfgang.
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL Patent: US 5663141-A 17 02-SEP-1997;
FEATURES Location/Qualifiers
source
BASE COUNT 21 a 21 c 17 g 9 t
ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 68;
Best Local Similarity 95.2%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32
||||| |||||||
DB 68 CCTATTACTGCAGGTATTCTT 48

RESULT 12
I63554/c
LOCUS I63554 68 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 17 from patent US 5663141.
ACCESSION I63554
VERSION I63554.1 GI:2481127
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68)
AUTHORS Kurfuerst,M., Ruebsamen,K., Schmied,B., Koerwer,W., Schweden,J. and
Hoeffken,H.Wolfgang.
TITLE Hirudin/polyalkylene glycol conjugates and hirudin muteins
JOURNAL Patent: US 5663141-A 17 02-SEP-1997;
FEATURES Location/Qualifiers
source
BASE COUNT 21 a 21 c 17 g 9 t
ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 68;
Best Local Similarity 95.2%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32
||||| |||||||
DB 68 CCTATTACTGCAGGTATTCTT 48

RESULT 13
A31392
LOCUS A31392 77 bp DNA PAT 03-NOV-1995
DEFINITION DNA fragment 3A from patent WO9108229.
ACCESSION A31392
VERSION A31392.1 GI:1247259
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 77)
AUTHORS
TITLE HIRUDINE POLYALKYLENE GLYCOL CONJUGATES
JOURNAL Patent: WO 9108229-A 9 13-JUN-1991;
FEATURES Location/Qualifiers
source
BASE COUNT 11 a 18 c 23 g 25 t
ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 77;
Best Local Similarity 95.2%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32
||||| |||||||
DB 5 CCTATTACTGCAGGTATTCTT 25

RESULT 14

```

I41184/c  
 LOCUS I41184  
 DEFINITION Sequence 4 from patent US 5624822.  
 ACCESSION I41184  
 VERSION I41184.1 GI:2081774  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 106)  
 AUTHORS Koerwer, W.  
 TITLE Hirudin fusion proteins and preparation of hirudin  
 JOURNAL Patent: US 5624822-A 4 29-APR-1997;  
 FEATURES  
 source  
 1..106  
 /organism="unknown"  
 BASE COUNT 36 a 28 c 28 g 14 t  
 ORIGIN

Query Match 60.6%; Score 19.4; DB 6; Length 106;  
 Best Local Similarity 95.2%; Pred. No. 2.3e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32  
 ||||| ||||| ||||| |||||  
 Db 106 CCTATTACTGCAGGTATTCTT 86

RESULT 15  
 SYNHIR/c  
 LOCUS SYNHIR 226 bp DNA linear SYN 27-APR-1993  
 DEFINITION Synthetic hirudin gene, complete cds.  
 ACCESSION M26762  
 VERSION M26762.1 GI:208478  
 KEYWORDS hirudin; proteinase inhibitor; thrombin inhibitor.  
 SOURCE Synthetic DNA.  
 ORGANISM  
 synthetic construct  
 artificial sequence.  
 1 (bases 1 to 226)  
 Bergmann, C., Dödt, J., Koehler, S., Fink, E. and Gassen, H.G.  
 Chemical synthesis and expression of a gene coding for hirudin, the  
 thrombin-specific inhibitor from the leech *Hirudo medicinalis*  
 Biol. Chem. Hoppe-Seyler 367, 731-740 (1986)  
 87026239  
 Location/Qualifiers  
 1..226  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 22..222  
 /note="hirudin"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AA72772.1"  
 /db\_xref="GI:208479"  
 /translation="MVYTDCTESQNLCLCEGSNVCGQGNKCILGSDGKNOCVTGE  
 GTPAPQSHNDGDFEIEPEYLQ"  
 63 a 56 c 61 g 46 t

BASE COUNT 63 a 56 c 61 g 46 t  
 ORIGIN

Query Match 60.6%; Score 19.4; DB 12; Length 226;  
 Best Local Similarity 95.2%; Pred. No. 2.4e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32  
 ||||| ||||| ||||| |||||  
 Db 226 CCTATTACTGCAGGTATTCTT 206

Search completed: July 15, 2002, 22:29:39  
 Job time: 19025 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:44:55 ; Search time 758.37 Seconds  
(without alignments)  
72.447 Million cell updates/sec

Title: US-10-053-641-8

Perfect score: 32

Sequence: 1 catcgccgcctatttcgaggtattctt 32

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	65.6	34	18 AAT49200	Hirudin gene linke
C 2	19.2	60.0	2085	23 ABL17165	Drosophila melanog
C 3	19.2	60.0	4466	23 ABL17164	Drosophila melanog
C 4	18.2	60.0	23379	23 AAS59510	Propionibacterium
5	19.2	60.0	349980	22 AAH41225	Pyrococcus abyss
6	19.2	60.0	349980	22 AAH41226	Pyrococcus abyss
C 7	19	59.4	1539	21 AA253344	Neisseria gonorrhe
8	18.8	58.8	45	16 AAQ88316	Hirudin gene (codo
9	18.8	58.8	45	16 AAQ85843	Primer to amplify

C 10	18.8	58.8	420	12 AAQ12155	Factor Xa-cleavabl
C 11	18.8	58.8	468	16 AAQ3244	Fusion construct o
12	18.8	58.8	737	22 AAH05529	Human cDNA clone (
13	18.8	58.8	877	20 AAH40024	Prostate cancer as
C 14	18.8	58.8	1305	23 AAS67097	DNA encoding novel
C 15	18.8	58.8	1467	12 AAQ12490	Factor Xa-cleavabl
C 16	18.4	57.5	726	12 AAQ12379	Hirudin peptide/Pr
C 17	18.4	57.5	2438	23 ABL23342	Drosophila melanog
C 18	18.2	56.9	743	22 AAT95116	Human neuroblastom
C 19	18.2	56.9	3203	11 AAQ06760	Sequence of beta-g
C 20	18.2	56.9	3900	23 ABL08940	Drosophila melanog
C 21	18.2	56.9	4324	23 ABL16366	Drosophila melanog
C 22	18.2	56.9	5059	21 AA236227	DNA encoding a per
C 23	18	56.2	37	19 AAV61849	A. contortrix prot
C 24	18	56.2	224	7 AAN60746	DNA encoding hirud
25	18	56.2	850	20 AAH40022	Prostate cancer as
26	18	56.2	2076	24 AAS62599	cDNA sequence #386
27	18	56.2	2215	20 AAH40015	Prostate cancer as
28	18	56.2	7995	24 ABK09796	Human ovarian tumo
29	18	56.2	8265	21 AAC75296	Human ORFX ORF851
30	18	56.2	8648	22 AAL26716	Human breast cance
31	18	56.2	8678	23 ABL27556	Drosophila melanog
32	18	56.2	15776	23 AAS59583	Propionibacterium
33	18	56.2	34878	22 AAK66167	Human immune/haema
34	18	56.2	34878	22 AAK80088	Human immune/haema
C 35	17.8	55.6	608	23 AAS91030	DNA encoding novel
C 36	17.8	55.6	667	22 AAF71326	Corynebacterium gl
C 37	17.8	55.6	1326	22 AAH68314	C glutamicum codin
C 38	17.8	55.6	1377	20 AAX27119	Calcineurin cataly
C 39	17.8	55.6	1634	22 AAS57011	C. trachomatis DNA
C 40	17.8	55.6	1767	17 AAT27106	Yeast calcineurin
C 41	17.8	55.6	2895	21 AA64753	C. trachomatis pmp
C 42	17.8	55.6	2895	22 AAS6256	Chlamydia trachoma
C 43	17.8	55.6	2934	21 AA64759	C. trachomatis pmp
C 44	17.8	55.6	2934	22 AAH56262	Chlamydia trachoma
C 45	17.8	55.6	2956	21 AA247010	Feline herpes viru

#### ALIGNMENTS

RESULT 1

AAT49200

ID AAT49200 standard; DNA; 34 BP.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX







```

AAH41226
ID  AAH41226 standard; DNA; 349980 BP.
XX
AC  AAH41226;
XX
XX  29-OCT-2001 (first entry)
XX
DE  Pyrococcus abyssi genomic fragment #5.
XX
XX  Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX
XX  Pyrococcus abyssi.
XX
XX  Key Location/Qualifiers
XX  misc_feature 1..49980
XX      /*tag= a
XX      /note= "This sequence overlaps with the 3' end of
XX      AAH41225"
XX  misc_feature 300001..349980
XX      /*tag= b
XX      /note= "This sequence overlaps with the 5' end of
XX      AAH41227"
XX
XX  FR2792651-A1.
XX
XX  27-OCT-2000.
XX
XX  21-APR-1999; 99FR-0005034.
XX
XX  21-APR-1999; 99FR-0005034.
XX
XX  (CNRS ) CNRS CENT NAT RECH SCI.
XX  (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX  Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX  Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
XX  WPI; 2001-126236/14.
XX
XX  New nucleotide sequences isolated from Pyrococcus abyssi encode
XX  proteins useful in industry -
XX
XX  Claim 1; Page 511-606; 1657pp; French.
XX
XX  The present invention relates to the genomic sequence of Pyrococcus
XX  abyssi and P. abyssi proteins (see AAB96053-AAB9842). P. abyssi is a
XX  hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
XX  vents. The present sequence is a fragment of the genomic sequence of P.
XX  abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41225
XX  and the 3' end of this sequence overlaps with the 5' end of AAH41227. The
XX  proteins of the present invention have various potential industrial uses,
XX  since the proteins are stable at very high temperatures, some up to 110
XX  degrees centigrade.
XX  Note: This patent is in the same patent family as WO200065062, which
XX  contains additional sequences as shown in AAB99132-AAB99143,
XX  AAH75903-AAH75920 and AAG66436.
XX
XX  Sequence 349980 BP; 98084 A; 80447 C; 77665 G; 93784 T; 0 other;

Query Match 60.0%; Score 19.2; DB 22; Length 349980;
Best Local Similarity 75.0%; Pred. No. 1.le+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 catgcggccgcctattattgcaggattctt 32
||||| ||||| ||||| ||||| |||||
DB 34771 catgagccctcccaataatgaaggtcactt 34802

RESULT 7
AAZ53344/C
ID  AAZ53344 standard; DNA; 1539 BP.
XX
XX
AC  AAZ53344;
XX
XX  21-MAR-2000 (first entry)
XX
XX  Neisseria gonorrhoeae ORF 155 partial DNA sequence SEQ ID NO:637.
XX
XX  Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX  antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX  antibacterial; gene therapy; ds.
XX
XX  Neisseria gonorrhoeae.
XX
XX  WO9957280-A2.
XX
XX  11-NOV-1999.
XX
XX  30-APR-1999; 99WO-US09346.
XX
XX  01-MAY-1998; 98US-0083758.
XX  31-JUL-1998; 98US-0094869.
XX  02-SEP-1998; 98US-0098994.
XX  02-SEP-1998; 98US-0099062.
XX  09-OCT-1998; 98US-0103749.
XX  09-OCT-1998; 98US-0103794.
XX  09-OCT-1998; 98US-0103796.
XX  25-FEB-1999; 99US-0121528.
XX
XX  (CHIR ) CHIRON CORP.
XX  (GENO-) INST GENOMIC RES.
XX
XX  Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX  Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX  Tettelin H, Venter JC;
XX
XX  WPI; 2000-062150/05.
XX  P-PSDB; AA74582.
XX
XX  Novel Neisserial polypeptides predicted to be useful antigens for
XX  vaccines and diagnostics -
XX
XX  Claim 7; Page 444; 1453pp; English.
XX
XX  AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX  represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX  and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
XX  PCR primers used in the exemplification of the present invention. The
XX  polypeptides, the polynucleotides, antibodies and compositions of
XX  the invention can be used as vaccines, as diagnostic reagents, and as
XX  immunogenic compositions. The polypeptides can be used in the
XX  manufacture of medicaments for treating or preventing infection due to
XX  Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX  presence of Neisseria bacteria, or to raise antibodies. They may also
XX  be used to screen for agonists or antagonists, which may themselves
XX  have use as antibacterial agents. The polynucleotides of the invention
XX  may also be used in gene therapy protocols.
XX
XX  Sequence 1539 BP; 330 A; 473 C; 428 G; 308 T; 0 other;

Query Match 59.4%; Score 19; DB 21; Length 1539;
Best Local Similarity 81.5%; Pred. No. 49;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 gcggccgcctattattgcaggatttc 30
||||| ||||| ||||| ||||| |||||
DB 669 GCTGCCCGCCGATCTTCAGGAAGTC 643

RESULT 8
AAQ88316
ID  AAQ88316 standard; DNA; 45 BP.
XX
XX  AAQ88316;

```



FT CDS /note= "encodes cleavage site IEGR"  
 FT 208..408  
 FT /\*tag= c  
 FT /label= hirudin gene  
 XX  
 PN W09109125-A.  
 XX  
 PD 27-JUN-1991.  
 XX  
 XX 07-DEC-1990; 90WO-GB01911.  
 XX  
 PF 07-DEC-1990; 90WO-GB01911.  
 PR  
 PR 07-DEC-1989; 89GB-0027722.  
 XX  
 XX (BRBI-) BRIT BIO-TECHN LTD.  
 PA  
 PI Dawson KM, Hunter MG, Czaplewski LG;  
 PI  
 XX WPI; 1991-208151/28.  
 DR P-PSDB; AAR12888.  
 XX  
 XX Fusion protein cleavage by blood clotting enzyme - for prodn. of  
 PT fractions having greater antithrombotic activity for therapy and  
 PT prophylaxis.  
 PT  
 XX Disclosure; Page 78; 115pp; English.  
 PS  
 XX The sequence of the synthetic hirudin HV-1 genes was designed  
 CC based on the published amino acid sequence (Dott J., et al FEBS  
 CC Letters 165 180 (1984)). Unique restriction sites were incorpor-  
 CC ated to facilitate subsequent genetic manipulation. The codons  
 CC selected were those favourable for *S. cerevisiae* or *E. coli*. The  
 CC sequence was divided into 12 oligomers which were synthesised and  
 CC then annealed. The ligation prod. was ligated to HindIII and EcoRI  
 CC treated pUC19 plasmid DNA and the resulting vector used to trans-  
 CC form *E. coli* K12 HB87. Plasmid pUC19 HV-1 was isolated from trans-  
 CC formants and inserted into plasmid pSW6, a shuttle vector, for ex-  
 CC pression. The gene was then used to construct an expression vector  
 CC in which two hirudin genes are linked together via a linker encod-  
 CC ing a cleavage site for factor Xa. The factor Xa is present at the  
 CC site of the target thrombus so the active agents are released  
 CC specifically at the place where clot formation is occurring.  
 CC See also AAQ12153-Q12156, AAQ12158-Q12162 and AAQ12490.  
 XX  
 XX Sequence 420 BP; 124 A; 90 C; 110 G; 96 T; 0 other;

Query Match 58.8%; Score 18.8; DB 12; Length 420;  
 Best Local Similarity 90.9%; Pred. No. 47;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 11 cccattattgcaggattctt 32  
 ||||| ||||| ||||| |||||  
 Db 410 CCTATTACTGCAGATATCTT 389

RESULT 11  
 AAQ93244/c  
 ID AAQ93244 standard; DNA; 468 BP.  
 XX  
 XX AAQ93244;  
 AC

DT 01-NOV-1995 (first entry)  
 XX

DE Fusion construct of glucoamylase-hirudin DNA.

KW expression cassette; recombinant protein; production; Hansenula;  
 KW yeast; leader sequence; adaptor; alpha helix; glucoamylase;  
 KW secretion; processing; thrombin inhibitor; hirudin; ds.  
 XX  
 XX Synthetic.

OS  
 XX  
 FH Key Location/Qualifiers

FT primer\_bind complement (1..33)  
 FT /\*tag= a  
 FT /note= "primer AAQ85840 binding site"  
 FT 13..444  
 FT CDS  
 FT /\*tag= b  
 FT /product= glucoamylase-hirudin fusion protein  
 FT 199..223  
 FT primer\_bind  
 FT /\*tag= c  
 FT /note= "primer AAQ85841 binding site"  
 FT complement (220..267)  
 FT /\*tag= d  
 FT /note= "primer AAQ85842 binding site"  
 FT 13..228  
 FT misc\_feature  
 FT /\*tag= e  
 FT /note= "encodes amino acids 1-72 of glucoamylase"  
 FT 241..246  
 FT misc\_feature  
 FT /\*tag= f  
 FT /note= "encodes processor signal"  
 FT 247..444  
 FT misc\_feature  
 FT /\*tag= g  
 FT /note= "encodes hirudin-h120"  
 FT 0..468  
 FT primer\_bind  
 FT /\*tag= h  
 FT /note= "primer AAQ85843 binding site"  
 FT  
 XX  
 XX DE4329969-A.  
 PN  
 XX  
 XX 09-MAR-1995.  
 PD  
 XX  
 XX 04-SEP-1993; 93DE-4329969.  
 PF  
 XX  
 XX 04-SEP-1993; 93DE-4329969.  
 PR  
 XX  
 XX (BADI ) BASF AG.  
 XX  
 XX (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.  
 PA  
 PI Bollschweiler C, Janowicz ZA, Piontek M, Schweden J;  
 PI Strasserawm, Weydemann U;  
 PI  
 XX WPI; 1995-107810/15.  
 DR P-PSDB; AAR76951.  
 DR  
 XX Recombinant protein prodn. in Hansenula yeast - transformed with  
 PT expression cassette contg. leader, adaptor, processing signal and  
 PT gene, provides efficient secretion and correct processing  
 PT  
 XX Example 1; Fig 1; 10pp; German.

XX AAQ93244 is a DNA fusion product resulting from ligation of an  
 CC EcoRI-PvuI fragment (see AAQ85840-41) and a PvuI-SalI fragment, encoding  
 CC a leader sequence, an adaptor (see AAR71472, amino acids 23-72 of GAM  
 CC (glucamylase from Schwanniomycetes occidentalis, plus -His-Pro-Leu-Gln at  
 CC the C-terminal end), a processor peptide (Lys-Arg) and a structural gene,  
 CC encoding hirudin. AAQ93244 is an example of an expression vector insert  
 CC of the general formula: L-A-P-Gene (L = leader sequence; A = adaptor;  
 CC P = processor; Gene = structural gene). The cassettes ensure efficient  
 CC secretion and correct processing of heterologous structural genes in  
 CC yeast of the genus *Hansenula*, and so provides high yields of mature  
 CC proteins and facilitates subsequent purification.  
 XX  
 XX Sequence 468 BP; 126 A; 110 C; 112 G; 120 T; 0 other;

Query Match 58.8%; Score 18.8; DB 16; Length 468;  
 Best Local Similarity 90.9%; Pred. No. 48;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 11 cccattattgcaggattctt 32  
 | ||||| ||||| |||||  
 Db 449 CTCTATTACTGCAGGTATCTT 428

RESULT 12

AAH05529  
ID AAH05529 standard; cDNA; 737 BP.  
XX AC AAH05529;  
XX DT 26-JUN-2001 (first entry)  
XX DE Human cDNA clone (5'-primer) SEQ ID NO:2364.  
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX OS Homo sapiens.  
XX PN EP1074617-A2.  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-0116126.  
XX PR 29-JUL-1999; 99JP-0248036.  
XX PR 27-AUG-1999; 99JP-0300253.  
XX PR 11-JAN-2000; 2000JP-0118776.  
XX PR 02-MAY-2000; 2000JP-0183767.  
XX PR 09-JUN-2000; 2000JP-0241899.  
XX PA (HELI-) HELIX RES INST.  
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.  
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs.  
XX PS Claim 1; SEQ ID 2364; 2537pp + CD ROM; English.  
XX CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification, where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX SQ Sequence 737 BP; 179 A; 206 C; 212 G; 136 T; 4 other;

Query Match 58.8%; Score 18.8; DB 22; Length 737;  
Best Local Similarity 76.7%; Pred. No. 53;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 3 tgcggccgcctattatgcagggtattctt 32  
||| ||||| | ||||| ||| |||  
Db 688 tgcggccgcctattatgcagggtattctt 717

RESULT 13  
AAH0024  
ID AAX40024 standard; DNA; 877 BP.  
XX AC AAX40024;  
XX DT 02-JUL-1999 (first entry)  
XX DE Prostate cancer associated gene.  
XX KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
XX OS Homo sapiens.  
XX PN WO9904265-A2.  
XX PD 28-JAN-1999.  
XX PF 15-JUL-1998; 98WO-US14679.  
XX PR 22-JUN-1998; 98US-0102322.  
XX PR 17-JUL-1997; 97US-0896164.  
XX PR 10-OCT-1997; 97US-0061599.  
XX PR 10-OCT-1997; 97US-0061765.  
XX PR 10-OCT-1997; 97US-0948705.  
XX PR 11-OCT-1997; 97GB-0021697.  
XX PA (LUDW-) LUDWIG INST CANCER RES.  
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;  
PI Tureci O;  
XX WPI: 1999-132448/11.  
XX PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
XX PS Claim 67; Page 638-639; 787pp; English.  
XX CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
XX SQ Sequence 877 BP; 210 A; 248 C; 250 G; 166 T; 3 other;

Query Match 58.8%; Score 18.8; DB 20; Length 877;  
Best Local Similarity 76.7%; Pred. No. 54;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 3 tgcggccgcctattatgcagggtattctt 32  
||| ||||| | ||||| ||| |||  
Db 680 tgcggccgcctattatgcagggtattctt 709

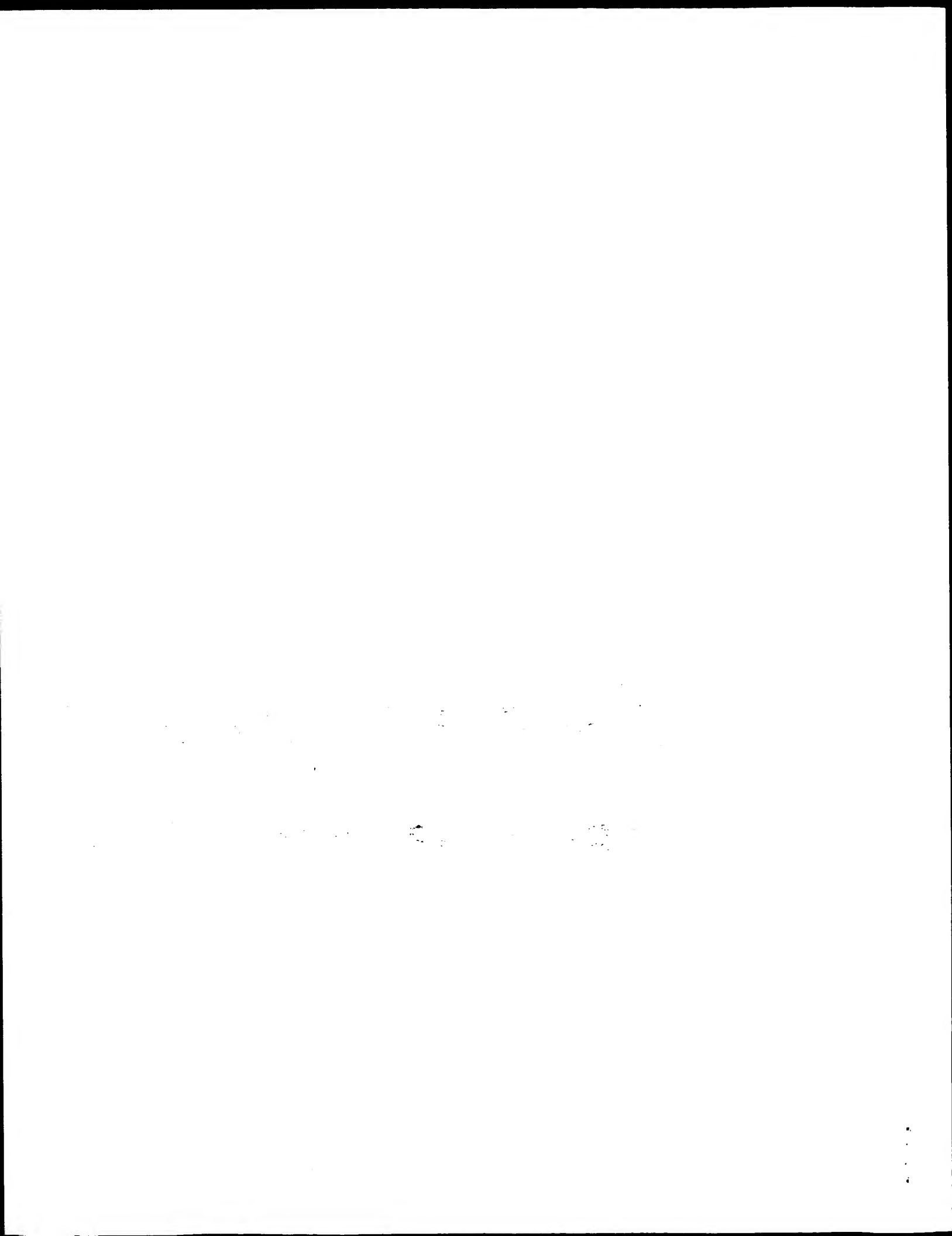
RESULT 14  
AAS67097/c

ID AA567097 standard; CDNA; 1305 BP.  
 XX AC  
 XX AA567097;  
 XX DT  
 XX 13-FEB-2002 (first entry)  
 XX DE  
 XX DNA encoding novel human diagnostic protein #2901.  
 XX KW  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX OS  
 XX Homo sapiens.  
 XX PN  
 XX WO200175067-A2.  
 XX PD  
 XX 11-OCT-2001.  
 XX PF  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX PR  
 XX 31-MAR-2000; 2000US-0540217.  
 XX PR  
 XX 23-AUG-2000; 2000US-0649167.  
 XX PA  
 XX (HYSE-) HYSEQ INC.  
 XX PI  
 XX Drmanac RT, Liu C, Tang YT;  
 XX DR  
 XX WPI; 2001-639362/73.  
 DR P-PSDB; AEG02910.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 1; SEQ ID No 2901; 103pp; English.  
 CC  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1305 BP; 345 A; 336 C; 357 G; 267 T; 0 other;  
  
 Query Match 58.8%; Score 18.8; DB 23; Length 1305;  
 Best Local Similarity 76.7%; Pred. No. 59;  
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
 QY 1 catgcggcgccctattattgcaggatttc 30  
 ||||| ||| ||||| |||  
 Db 1248 CATGGGGTCCACATTGTTGCTGATGTC 1219  
  
 RESULT 15  
 AAQ12490/c  
 ID AAQ12490 standard; DNA; 1467 BP.

XX AAQ12490;  
 AC  
 XX 17-SEP-1991 (first entry)  
 DT  
 XX  
 DE  
 XX Factor Xa-cleavable streptokinase-IEGR-hirudin gene.  
 DE  
 XX  
 KW Fusion protein; blood clotting; coagulation; fibrinolysis;  
 KW antithrombotic; thrombolysis; ss.  
 XX OS  
 XX Synthetic.  
 XX FH  
 XX Key  
 FT CDS  
 FT 1..1455  
 FT /\*tag= a  
 FT mat\_peptide  
 FT 1..1242  
 FT /\*tag= b  
 FT /label= streptokinase  
 FT 1243..1254  
 FT misc\_RNA  
 FT /\*tag= c  
 FT /label= linker  
 FT /note= encodes factor Xa cleavage site"  
 FT 1255..1453  
 FT mat\_peptide  
 FT /\*tag= d  
 FT /label= hirudin HV-1  
 FT  
 XX  
 PN WO9109125-A.  
 XX  
 XX 27-JUN-1991.  
 PD  
 XX  
 XX 07-DEC-1990; 90WO-GB01911.  
 XX  
 XX 07-DEC-1990; 90WO-GB01911.  
 PR  
 XX 07-DEC-1989; 89GB-0027722.  
 PR  
 XX  
 XX (BRBI-) BRIT BIO-TECHN LTD.  
 PA  
 XX  
 XX Dawson KM, Hunter MG, Czaplewski LG;  
 PI  
 XX WPI; 1991-208151/28.  
 DR P-PSDB; AAR12522.  
 DR  
 XX  
 XX Fusion protein cleavage by blood clotting enzyme - for prodn. of  
 PT fractions having greater antithrombotic activity for therapy and  
 PT prophylaxis.  
 PT  
 XX  
 PS Disclosure; Page 98; 115pp; English.  
 CC  
 CC The sequence of the synthetic hirudin HV-1 gene was designed  
 CC based on the published amino acid sequence (Dodd J., et al FEBS  
 CC Letters 165 180 (1984)). The sequence of streptokinase was obtd.  
 CC from PCR amplified chromosomal DNA from S. equisimilis ATCC 10009  
 CC or ATCC 9642. The primers used for the PCR were based on the pub-  
 CC lished DNA sequence of S. equisimilis strain H46A (Malke, H., Roe,  
 CC B., and Ferretti, J.J., Gene 34 357-362 [1985]). The two  
 CC sequences were used to construct an expression vector in which the  
 CC streptokinase gene is linked to the hirudin gene via a linking  
 CC sequence encoding a cleavage site for factor Xa. The factor Xa is  
 CC present at the site of the target thrombus so the active agents are  
 CC released specifically at the place where clot formation is occurring.  
 CC See also AAQ12153-Q12156 and AAQ12158-Q12162.  
 XX  
 SQ Sequence 1467 BP; 494 A; 317 C; 292 G; 364 T; 0 other;

Query Match 58.8%; Score 18.8; DB 12; Length 1467;  
 Best Local Similarity 90.9%; Pred. No. 60;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
 QY 11 ccctattattgcaggatttc 32  
 ||||| ||||| |||||  
 Db 1457 CCCTATTACTGCAGATATTCTT 1436

Search completed: July 15, 2002, 22:45:19  
Job time: 10092 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 22:31:52 ; Search time 174.99 seconds  
(without alignments)  
44.918 Million cell updates/sec

Title: US-10-053-641-8

Perfect score: 32

Sequence: 1 catgcggccgcctattattgcaggtattctt 32

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	65.6	34	1	US-08-466-265-13
2	20.8	65.0	128	1	US-08-262-384A-5
3	20.8	65.0	235	1	US-08-262-384A-8
C 4	19.4	60.6	68	1	US-07-859-453E-13
C 5	19.4	60.6	68	1	US-07-859-453E-15
C 6	19.4	60.6	68	1	US-07-859-453E-17
C 7	19.4	60.6	105	1	US-08-262-384A-4
C 8	19.4	60.6	235	1	US-08-262-384A-7
C 9	18.8	58.8	31	1	US-07-854-596B-60
C 10	18.8	58.8	45	1	US-08-619-598-8
C 11	18.8	58.8	45	1	US-08-605-053-5
C 12	18.8	58.8	223	1	US-07-854-596B-7
C 13	18.8	58.8	420	1	US-07-854-596B-8
C 14	18.8	58.8	1467	1	US-07-854-596B-46
C 15	18.4	57.5	726	1	US-08-262-384A-9
C 16	18.4	57.5	726	1	US-08-262-384A-10
C 17	17.8	55.6	1767	1	US-08-328-322-20
C 18	17.8	55.6	2353	1	US-08-328-322-11
C 19	17.6	55.0	40	2	US-08-839-581A-25
C 20	17.6	55.0	40	4	US-09-023-591A-25
C 21	17.6	55.0	8491	2	US-08-757-439-1
C 22	17.4	54.4	3157	6	5198347-3
C 23	17	53.1	30	1	US-08-093-741-21
C 24	17	53.1	30	1	US-08-720-012-21
C 25	17	53.1	30	2	US-08-560-098A-26
C 26	17	53.1	30	3	US-08-967-024C-14
C 27	17	53.1	50	1	US-08-093-741-24

#### ALIGNMENTS

##### RESULT 1

US-08-466-265-13

; Sequence 13, Application US/08466265

; Patent No. 5712114

; GENERAL INFORMATION:

; APPLICANT: Mankovich, John A.

; TITLE OF INVENTION: Compositions for Expression of Proteins in Host

; TITLE OF INVENTION: Cells Using a Preprocollagen Signal Sequence

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 60 State Street, suite 510

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/466,265

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: DeConti, Giulio A., Jr.

; REGISTRATION NUMBER: 31,503

; REFERENCE/DOCKET NUMBER: BBI-030

; TELEPHONE: (617)227-7400

; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 34 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: oligonucleotide

; US-08-466-265-13

Query Match 65.6%; Score 21; DB 1; Length 34;

Best Local Similarity 82.8%; Pred. No. 0.42;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 gcggccgcctattattgcaggtattctt 32

||||| ||||||| ||||||| |||||||

Db 1 GGGGGATCCCTATTACTGCGAGGTATTCTT 29



Query Match 60.6%; Score 19.4; DB 1; Length 68;  
 Best Local Similarity 95.2%; Pred. No. 2.6;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32  
 Db 68 CCTATTACTGCAGGTATTCTT 48

## RESULT 5

US-07-859-453E-15/c  
 ; Sequence 15, Application US/07859453E  
 ; Patent No. 5663141  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kurfuerst, Manfred  
 ; APPLICANT: Ruebsamen, Klaus  
 ; APPLICANT: Schmied, Bernhard  
 ; APPLICANT: Koerwer, Wolfgang  
 ; APPLICANT: Schweden, Juergen  
 ; APPLICANT: Hoeftken, Hans Wolfgang  
 ; TITLE OF INVENTION: Hitudin/polyalkylene glycol  
 ; NUMBER OF SEQUENCES: 28  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kell & Weinkauff  
 ; STREET: 1101 Connecticut Avenue  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
 ; OPERATING SYSTEM: MS-DOS version 7.0  
 ; SOFTWARE: WordPerfect version 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/859,453E  
 ; FILING DATE: 29-MAY-1992  
 ; CLASSIFICATION: A 61 K 37/64  
 ; CLASSIFICATION: A 61 K 47/48  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/EP90/01998  
 ; FILING DATE: 22-NOV-1990  
 ; INFORMATION FOR SEQ ID NO: 15:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 68 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; US-07-859-453E-15

Query Match 60.6%; Score 19.4; DB 1; Length 68;  
 Best Local Similarity 95.2%; Pred. No. 2.6;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32  
 Db 68 CCTATTACTGCAGGTATTCTT 48

## RESULT 6

US-07-859-453E-17/c  
 ; Sequence 17, Application US/07859453E  
 ; Patent No. 5663141  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kurfuerst, Manfred  
 ; APPLICANT: Ruebsamen, Klaus  
 ; APPLICANT: Schmied, Bernhard  
 ; APPLICANT: Koerwer, Wolfgang

; APPLICANT: Schweden, Juergen  
 ; APPLICANT: Hoeftken, Hans Wolfgang  
 ; TITLE OF INVENTION: Hitudin/polyalkylene glycol  
 ; NUMBER OF SEQUENCES: 28  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kell & Weinkauff  
 ; STREET: 1101 Connecticut Avenue  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
 ; OPERATING SYSTEM: MS-DOS version 7.0  
 ; SOFTWARE: WordPerfect version 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/859,453E  
 ; FILING DATE: 29-MAY-1992  
 ; CLASSIFICATION: 514  
 ; CLASSIFICATION: A 61 K 37/64  
 ; CLASSIFICATION: A 61 K 47/48  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/EP90/01998  
 ; FILING DATE: 22-NOV-1990  
 ; INFORMATION FOR SEQ ID NO: 17:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 68 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; US-07-859-453E-17

Query Match 60.6%; Score 19.4; DB 1; Length 68;  
 Best Local Similarity 95.2%; Pred. No. 2.6;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattattgcagggtattctt 32  
 Db 68 CCTATTACTGCAGGTATTCTT 48

## RESULT 7

US-08-262-384A-4/c  
 ; Sequence 4, Application US/08262384A  
 ; Patent No. 5624822  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Koerwer, Wolfgang  
 ; TITLE OF INVENTION: The Preparation of Hitudin  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kell & Weinkauff  
 ; STREET: 1101 Connecticut Avenue  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
 ; OPERATING SYSTEM: MS-DOS version 6.0  
 ; SOFTWARE: WordPerfect version 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/262,384A  
 ; FILING DATE: 20-JUN-1994  
 ; CLASSIFICATION: 530  
 ; CLASSIFICATION: C 12 N 15/62  
 ; CLASSIFICATION: C 12 N 15/31  
 ; CLASSIFICATION: C 07 K 7/10  
 ; CLASSIFICATION: C 12 P 21/02  
 ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990  
APPLICATION NUMBER: US 07861820  
FILING DATE: 18-JUN-1992  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-262-384A-4

Query Match 60.6%; Score 19.4; DB 1; Length 106;  
Best Local Similarity 95.2%; Pred. No. 2.9;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattatgcagggtattctt 32  
||||| |||||||  
DB 106 CCTATTACTGCAGGTATTCTT 86

RESULT 8  
US-08-262-384A-7/c  
Sequence 7, Application US/08262384A  
Patent No. 5624822  
GENERAL INFORMATION:  
APPLICANT: Koerwer, Wolfgang  
TITLE OF INVENTION: The Preparation of Hirudin  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM AT-compatible, 80486 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/262,384A  
FILING DATE: 20-JUN-1994  
CLASSIFICATION: 530  
CLASSIFICATION: C 12 N 15/62  
CLASSIFICATION: C 12 N 15/31  
CLASSIFICATION: C 07 K 7/10  
CLASSIFICATION: C 12 P 21/02  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/02084  
FILING DATE: 04-DEC-1990  
APPLICATION NUMBER: US 07861820  
FILING DATE: 18-JUN-1992  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-262-384A-7

Query Match 60.6%; Score 19.4; DB 1; Length 235;  
Best Local Similarity 95.2%; Pred. No. 3.4;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 cctattatgcagggtattctt 32  
||||| |||||||  
DB 235 CCTATTACTGCAGGTATTCTT 215

RESULT 9  
US-07-854-596B-60/c  
Sequence 60, Application US/07854596B  
Patent No. 5434073  
GENERAL INFORMATION:  
APPLICANT: Dawson, Keith M  
APPLICANT: Hunter, Michael G  
APPLICANT: Czaplewski, Lloyd G  
TITLE OF INVENTION: Proteins and nucleic acids  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dr. John J. McDonnell  
STREET: Ten South Wacker Drive, Suite 3000  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/854,596B  
FILING DATE: 03-JUN-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: McDonnell, John J  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 92,337  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
TELEX: 910-221-5317  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..31  
OTHER INFORMATION: /note= "oligonucleotide BB2021"  
US-07-854-596B-60

Query Match 58.8%; Score 18.8; DB 1; Length 31;  
Best Local Similarity 90.9%; Pred. No. 4.2;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 11 cctattatgcagggtattctt 32  
||||| |||||||  
DB 26 CCTATTACTGCAGGTATTCTT 5

RESULT 10  
US-08-619-598-8  
Sequence 8, Application US/08619598  
Patent No. 5672487  
GENERAL INFORMATION:  
APPLICANT: SCHWEDEN, Juergen  
APPLICANT: BOLLSCHWEILER, Claus  
APPLICANT: PIONTEK, Michael  
APPLICANT: WEYDEMANN, Ulrike  
APPLICANT: GELLISSEN, Gerd  
TITLE OF INVENTION: THE RECOMBINANT PRODUCTION  
OF PROTEINS IN YEAST  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue



Best Local Similarity 90.9%; Pred. No. 6.3; Indels 2; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

Qy 11 ccctattattgcaggtattctt 32  
||||||| ||||| |||||  
Db 219 CCCTATTACTGCAGATATTCTT 198

## RESULT 13

US-07-854-596B-8/C  
; Sequence 8, Application US/07854596B  
; Patent No. 5434073  
; GENERAL INFORMATION:  
; APPLICANT: Dawson, Keith M  
; APPLICANT: Hunter, Michael G  
; APPLICANT: Czaplowski, Lloyd G  
; TITLE OF INVENTION: Proteins and nucleic acids  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dr. John J. McDonnell  
; STREET: Ten South Wacker Drive, Suite 3000  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/854,596B  
; FILING DATE: 03-JUN-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 92,337  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; TELEX: 910-221-5317  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 420 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..420  
; OTHER INFORMATION: /note= "Factor Xa-cleavable  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..402  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 1..402  
US-07-854-596B-8

Query Match 58.8%; Score 18.8; DB 1; Length 420;  
Best Local Similarity 90.9%; Pred. No. 7.1;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 11 ccctattattgcaggtattctt 32  
||||||| ||||| |||||  
Db 410 CCCTATTACTGCAGATATTCTT 389

## RESULT 14

US-07-854-596B-46/C  
; Sequence 46, Application US/07854596B  
; Patent No. 5434073  
; GENERAL INFORMATION:  
; APPLICANT: Dawson, Keith M  
; APPLICANT: Hunter, Michael G  
; APPLICANT: Czaplowski, Lloyd G  
; TITLE OF INVENTION: Proteins and nucleic acids  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dr. John J. McDonnell  
; STREET: Ten South Wacker Drive, Suite 3000  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/854,596B  
; FILING DATE: 03-JUN-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 92,337  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; TELEX: 910-221-5317  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1467 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..1467  
; OTHER INFORMATION: /note= "Streptokinase-hirudin  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1449  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 1..1449  
US-07-854-596B-46

Query Match 58.8%; Score 18.8; DB 1; Length 1467;  
Best Local Similarity 90.9%; Pred. No. 9.2; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 2;

Qy 11 ccctattattgcaggtattctt 32  
||||||| ||||| |||||  
Db 1457 CCCTATTACTGCAGATATTCTT 1436

## RESULT 15

US-08-262-384A-9/c  
; Sequence 9, Application US/08262384A  
; Patent No. 5624822  
; GENERAL INFORMATION:  
; APPLICANT: Koerwer, Wolfgang  
; TITLE OF INVENTION: The Preparation of Hirudin  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinkauff

```

; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262.384A
; FILING DATE: 20-JUN-1994
; CLASSIFICATION: 530
; CLASSIFICATION: C 12 N 15/62
; CLASSIFICATION: C 12 N 15/31
; CLASSIFICATION: C 07 K 7/10
; CLASSIFICATION: C 12 P 21/02
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02084
; FILING DATE: 04-DEC-1990
; APPLICATION NUMBER: US 07861820
; FILING DATE: 18-JUN-1992
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-262-384A-9

```

```

Query Match      57.5%   Score 18.4; DB 1; Length 726;
Best Local Similarity 95.0%   Pred. No. 12;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

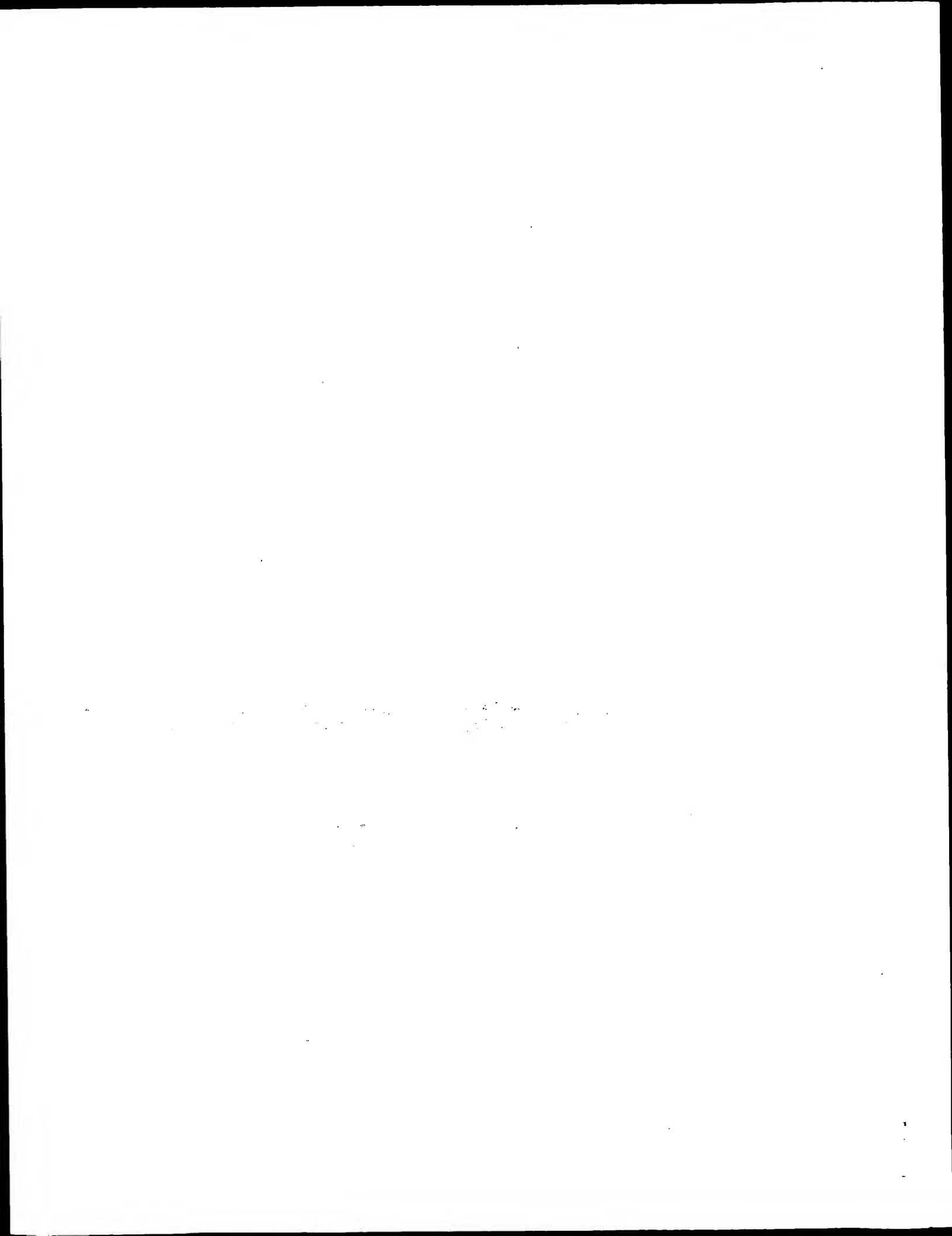
Qy 13 ctattattgcagggtattctt 32
Db 726 CTATTACTGCAGGTATTCTT 707

```

```

Search completed: July 15, 2002, 22:31:53
Job time: 18479 sec

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 21:49:22 ; Search time 6165.88 seconds  
(without alignments)  
70.047 Million cell updates/sec

Title: US-10-053-641-8

Perfect score: 32

Sequence: 1 catcgccgcgcctatttgcaggatttttt 32

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hcc:\*
- 9: gb\_estl:\*
- 10: gb\_est2:\*
- 11: gb\_hcc:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	65.6	1643	10	BG242109 602354656
2	20.8	65.0	737	10	BG781087 SEAMC001
3	20.4	63.7	329	10	BG652278
4	20.4	63.7	509	9	AW317727 sg56e07.y
5	20.4	63.7	542	10	BM143360
6	20.4	63.7	736	10	BG657243
7	20.4	63.7	1964	10	BF140681
8	20	62.5	1085	10	BI838572
9	19.8	61.9	525	10	BM180243
10	19.8	61.9	593	9	AL628245
11	19.8	61.9	655	9	AL595404
12	19.8	61.9	833	9	BE054965
13	19.8	61.9	1181	10	BE892107
14	19.6	61.3	952	12	AG098165
15	19.6	61.3	959	10	BG387896
16	19.6	61.3	1270	12	AG035742
17	19.4	60.6	258	10	BG762709 602734617

18	19.4	60.6	438	10	BG359980
19	19.4	60.6	480	9	AW761595
20	19.4	60.6	488	10	BI471822
21	19.4	60.6	646	10	BG599576
22	19.4	60.6	666	10	W14288
23	19.4	60.6	685	12	AG074640
24	19.4	60.6	726	12	AQ048734
25	19.4	60.6	772	12	AG069748
26	19.4	60.6	1007	10	BG176216
27	19.2	60.0	403	9	AI130633
28	19.2	60.0	409	12	BH618464
29	19.2	60.0	464	12	B25714
30	19.2	60.0	521	12	AZ586077
31	19.2	60.0	544	9	AI313785
32	19.2	60.0	796	10	BG193247
33	19.2	60.0	867	12	CNS0101E
34	19.2	60.0	898	10	BF104501
35	19.2	60.0	925	12	CNS03V97
36	19.2	60.0	978	10	BF234403
37	19	59.4	549	12	BH350672
38	19	59.4	556	12	BH188630
39	19	59.4	556	12	CNS07SLR
40	19	59.4	669	12	AG074692
41	19	59.4	692	12	AG125024
42	19	59.4	723	9	AV704699
43	18.8	58.8	212	9	AV382007
44	18.8	58.8	439	10	BG088443
45	18.8	58.8	483	12	AQ661602

#### ALIGNMENTS

RESULT 1  
BG242109 1643 bp mRNA linear EST 13-FEB-2001  
LOCUS 602354656F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:4483069 5',  
DEFINITION mRNA sequence.  
ACCESSION BG242109.1 GI:12751924  
VERSION BG242109  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 1643)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cyaphs@email.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10320 row: p column: 14  
High quality sequence stop: 34.  
Location/Qualifiers  
1. 1643  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4483069"  
/clone\_lib="NCI\_CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"  
/lab\_host="DH10B"  
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator

```

BASE COUNT      478 a 525 c 283 g 357 t
ORIGIN
    providing samples: Gilbert Smith, NIH"

Query Match      65.6%; Score 21; DB 10; Length 1643;
Best Local Similarity 82.8%; Pred. No. 2e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 ggcggcgccctattatgcaggattctt 32
    ||||| ||||| ||||| ||||| |||||
Db 25 GGGCGCGCCCTATTATTTGTATATATT 53

RESULT 2
BG781087      737 bp mRNA linear EST 20-MAY-2001
LOCUS      SEAMC001044 Sea urchin primary mesenchyme cell cDNA library
DEFINITION Strongylocentrotus purpuratus cDNA clone PC_0011_A2_H05_MR 5', mRNA
            sequence.
ACCESSION   BG781087
VERSION     BG781087.1 GI:14152100
KEYWORDS    EST.
SOURCE      Strongylocentrotus purpuratus.
ORGANISM    Strongylocentrotus purpuratus.
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoidea;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1 (bases 1 to 737)
AUTHORS     Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
            Ettensohn,C.A.
TITLE       A large scale analysis of mRNAs expressed by primary mesenchyme
            cells of the sea urchin embryo
JOURNAL     Development 128 (13), 2615-2627 (2001)
MEDLINE     21384984
COMMENT     Contact: Ettensohn CA
            Dept. Biol. Sci.
            Carnegie Mellon University
            4400 Fifth Avenue, Pittsburgh, PA 15213, USA
            Tel: +1 412 268 5849
            Email: ettensohn@andrew.cmu.edu.

FEATURES
    source
        Location/Qualifiers
            1..737
                /organism="Strongylocentrotus purpuratus"
                /db_xref="taxon:7668"
                /clone="PC_0011_A2_H05_MR"
                /clone_lib="Sea urchin primary mesenchyme cell cDNA
                library"
                /tissue_type="embryo"
                /cell_type="primary mesenchyme cells"
                /lab_host="E.coli"
                /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
                dt priming from poly A+ RNA, directionally cloned"
BASE COUNT      269 a 110 c 122 g 236 t
ORIGIN
    Query Match      65.0%; Score 20.8; DB 10; Length 737;
    Best Local Similarity 78.1%; Pred. No. 2e+02;
    Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 catgcggcgccctattatgcaggattctt 32
    ||||| ||||| ||||| ||||| |||||
Db 32 CACGCGTCCCATATTATTTGTATATATTCTT 63

RESULT 3
BG652278      329 bp mRNA linear EST 29-NOV-2001
LOCUS      sad64h03.y1 Gm-cl051 Glycine max cDNA clone GENOME SYSTEMS CLONE
DEFINITION ID: Gm-cl051-5333 5', mRNA sequence.
ACCESSION   BG652278
VERSION     BG652278.1 GI:13789687

BASE COUNT      478 a 525 c 283 g 357 t
ORIGIN
    providing samples: Gilbert Smith, NIH"

KEYWORDS
SOURCE
ORGANISM
EST.
    soybean.
    Glycine max
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
    Glycine.
    1 (bases 1 to 329)
    Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Corryell,V., Khanna
    ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
    Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
    ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
    ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
    ,R., Waterston,R. and Willson,R.
    Public Soybean EST Project
    Unpublished (1999)
    Contact: Shoemaker R/Public Soybean EST Project
    Public Soybean EST Project
    Washington University School of Medicine
    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
    Tel: 314 286 1800
    Fax: 314 286 1810
    Email: est@watson.wustl.edu
    This clone is available through: ResGen, Invitrogen Corp. 2130
    South Memorial Parkway Huntsville, AL 35801 For further information
    call: (800) 533-4363 or contact via email: ccu@resgen.com
    High quality sequence stop: 316.
    Location/Qualifiers
        1..329
            /organism="Glycine max"
            /db_xref="taxon:3847"
            /clone="GENOME SYSTEMS CLONE ID: Gm-cl051-5333"
            /clone_lib="Gm-cl051"
            /tissue_type="floral meristematic mRNA"
            /lab_host="DH10B"
            /note="vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
            XhoI; The cDNA library was constructed from floral
            meristematic mRNA provided by Dr. Halina Knap of Clemson
            University. Complementary DNA was synthesized from mRNA
            using a primer consisting of a poly(dT) sequence with a
            XhoI restriction site. EcoRI adapters were ligated to the
            blunt-ended cDNA fragments followed by XhoI digestion. The
            cDNA fragments were directionally cloned into the
            EcoRI-XhoI restriction site of the pBluescript vector. The
            ligated cDNA fragments were transformed into DH10B host
            cells (GibcoBRL). This library was constructed in the
            laboratory of Dr. Randy Shoemaker."
BASE COUNT      90 a 59 c 58 g 122 t
ORIGIN
    Query Match      63.7%; Score 20.4; DB 10; Length 329;
    Best Local Similarity 80.0%; Pred. No. 2.5e+02;
    Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 tgcggcgccctattatgcaggattctt 32
    ||||| ||||| ||||| ||||| |||||
Db 206 TGCTGCTGCCATATCATTCGAGGTAATTTT 235

RESULT 4
AW317727
LOCUS      sg56e07.y1 Gm-cl1007 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl1007-445 5', mRNA sequence.
ACCESSION   AW317727
VERSION     AW317727.1 GI:6747271
KEYWORDS    soybean.
            Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            EST.
            SOURCE
            ORGANISM
            soybean.
            Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

```

REFERENCE  
AUTHORS  
1 (bases 1 to 542)  
Shoemaker, R., Kelm, P., Vodkin, L., Expelding, J., Coryell, V., Khanna  
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.

RESULT 6  
E657243/C  
LOCUS  
DEFINITION  
ACCESSION  
VERSION

OCUS BE657243  
DEFINITION GM700001A20C8 Gm-r1070 Glycine max cDNA clone Gm-r1070-112 3', mRNA linear EST 24-MAY-2001  
736 bp



CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL).  
 DNA Sequencing by: InCyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM11566 row: a column: 14  
 High quality sequence stop: 601.  
 Location/Qualifiers  
 1..1085

## FEATURES

source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5224981"  
 /clone\_lib="NIH\_MGC\_120"  
 /lab\_host="DH10B"  
 /note="Organ: pooled pancreas and spleen; Vector:  
 pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA  
 source anonymous pool of spleen and pancreas from 28 yo  
 male. Library is oligo-dr primed and directionally cloned  
 (EcoRV site is destroyed upon cloning). Average insert  
 size 1.5 kb, insert size range 1-2.5 kb. Library is  
 normalized and enriched for full-length clones and was  
 constructed by C. Gruber (Invitrogen). Research Genetics  
 tracking code 025. Note: this is a NIH\_MGC Library."

332 a 306 c 338 g 109 t  
 BASE COUNT  
 ORIGIN

Query Match 62.5%; Score 20; DB 10; Length 1085;  
 Best Local Similarity 82.1%; Pred. No. 4.7e+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 ggcggcgccctatttcgaggtattct 31  
 ||||| ||||| || ||||| |||||  
 Db 540 GCGCGCGCCCTCTTTTGCAGCTTTCT 513

## RESULT 9

BM180243  
 LOCUS BM180243 525 bp mRNA linear EST 07-DEC-2001  
 DEFINITION daJ87b06.x1 NICHD XGC L11 Xenopus laevis cDNA clone IMAGE:5129795  
 3', mRNA sequence.

ACCESSION BM180243  
 VERSION BM180243.1 GI:17404314  
 KEYWORDS EST.  
 SOURCE African clawed frog.

## ORGANISM

Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodinae; Xenopus.

1 (bases 1 to 525)

NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgap@rmail.nih.gov](mailto:cgap@rmail.nih.gov)

Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: Xenopus clones from this library are available  
 through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Seq primer: -40UP from Gibco

High quality sequence stop: 427.

## FEATURES

source  
 1..525  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:5129795"  
 /clone\_lib="NICHD XGC L11"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dr.

Average insert size 1.4 kb. Constructed by Life  
 Technologies. Note: This is a Xenopus Gene Collection (XGC  
 ) library."

BASE COUNT 170 a 77 c 101 g 177 t  
 ORIGIN

Query Match 61.9%; Score 19.8; DB 10; Length 525;  
 Best Local Similarity 77.4%; Pred. No. 4.8e+02;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 atcgggcgccctatttcgaggtattctt 32  
 ||||| ||||| ||||| ||||| |||||  
 Db 97 ATGCCCGCTGTGTAFTATTCATGATTCCT 127

## RESULT 10

AL628245  
 LOCUS AL628245 593 bp mRNA linear EST 02-NOV-2001  
 DEFINITION AL628245 XGC-gastrula silurana tropicalis cDNA clone TGas009a06 5',  
 mRNA sequence.

ACCESSION AL628245  
 VERSION AL628245.1 GI:16597728  
 KEYWORDS EST.  
 SOURCE western clawed frog.

## ORGANISM

Silurana tropicalis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodinae; Silurana.

1 (bases 1 to 593)

Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (10\_2001)

Unpublished (2001)

Contact: Huckle E

Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: [trp@sanger.ac.uk](mailto:trp@sanger.ac.uk)

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TGas009a06.sp6

Sequencing primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

## FEATURES

source  
 1..593  
 /organism="Silurana tropicalis"  
 /db\_xref="taxon:8364"  
 /clone="TGas009a06"

/clone\_lib="XGC-gastrula"

/dev\_stage="gastrula (stages 10.5-13 mixed)"

/lab\_host="Escherichia coli XL1-blue"

/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; CDNA

was oligo dr primed from 5ug of poly A+ RNA from stages

10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated

into pCS107 with EcoRI at the 5' end and NotI at the 3'

end."

BASE COUNT 143 a 135 c 169 g 146 t

## ORIGIN

Query Match 61.9%; Score 19.8; DB 9; Length 593;  
 Best Local Similarity 77.4%; Pred. No. 5e+02;  
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 catcgccgcccctatttcgaggtattct 31

|| ||||| ||||| ||||| |||||

Db 549 CAAGCGGGGCCCTATTCCTGCAGTCTCT 579

## RESULT 11

AL595404  
 LOCUS AL595404 655 bp mRNA linear EST 30-JUL-2001  
 DEFINITION AL595404 XGC-gastrula silurana tropicalis cDNA clone TGas006g19 5',  
 mRNA sequence.

ACCESSION	AL595404	1	GI:15007479	
VERSION	EST.			
KEYWORDS	western clawed frog.			
SOURCE	Silurana tropicalis			
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Silurana.			
REFERENCE	1 (bases 1 to 655)			
AUTHORS	Huckle,E., Taylor,R., McMurray,A., Ashurst,J.L., Zorn,A.M. and Rogers,J.			
TITLE	Sanger xenopus tropicalis EST project 2001			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Huckle E Sanger Centre Hinxton, Cambridgeshire, CB10 1SA, UK Email: trop@sanger.ac.uk Sanger xenopus tropicalis EST project 2001 TROPICALIS_SEQUENCE_ID: TGas006g19.sp6 Sequencing primer: SP6 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.			
FEATURES	Location/Qualifiers			
source	1. 655 /organism="Silurana tropicalis" /db_xref="taxon:8364" /clone="TGas006g19" /clone_lib="XGC-gastrula" /dev_stage="gastrula (stages 10.5-13 mixed)" /lab_host="Escherichia coli XL1-blue" /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."			
BASE COUNT	169 a 148 c 180 g 158 t			
ORIGIN				
Query Match	61.9%; Score 19.8; DB 9; Length 655;			
Best Local Similarity	77.4%; Pred. No. 5.1e+02;			
Matches	24; Conservative 0; Mismatches 7; Indels 0; Gaps			
QY	1 catgcgcgcgcctattatgcagggtattct 31 			
Db	579 CAAGCGGGGGCCCTATTCTGCAGTCTTCT 609			
RESULT	12			
BE054965				
LOCUS	BE054965 833 bp mRNA linear EST 07-MAR-2001			
DEFINITION	GA_Ea0032D15f Gossypium arboreum 7-10 dpa fiber library Gossypium			
ACCESSION	BE054965			
VERSION	BE054965			
KEYWORDS	arboreum cDNA clone GA_Ea0032D15f, mRNA sequence.			
SOURCE	BE054965.2 GI:13245904			
ORGANISM	EST.			
REFERENCE	Gossypium arboreum.			
AUTHORS	Gossypium arboreum			
TITLE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustoids II; Malvales; Malvaceae; Gossypium.			
JOURNAL	1 (bases 1 to 833)			
COMMENT	Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry ,D., Wood,T.C., Leslie,A. and Wilkins,T.A. An integrated analysis of the genetics, development, and evolution of the cotton fiber Unpublished (2000) On Jun 8, 2000 this sequence version replaced gi:8382022. Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288			



